RL-TR-94-20, Vol I (of two)
In-House Report
February 1994

3

AD-A277 202

PROCEEDINGS OF THE 1993 ANTENNA APPLICATIONS SYMPOSIUM

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-94-20, Vol I (of two) has been reviewed and is approved for publication.

APPROVED:

DANIEL J. JACAVANCO
Chief, Antennas and Components Division
Electromagnetics and Reliability Directorate

FOR THE COMMANDER: John K Schindler

JOHN K. SCHINDLER
Director of Electromagnetics & Reliability

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL (ERAS) Hanscom AFB MA 01731. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE February 1994	3. REPORT TYPE AND DATES COVERED In-House Report

4. TITLE AND SUBTITLE

PROCEEDINGS OF THE 1993 ANTENNA APPLICATIONS SYMPOSIUM

5. FUNDING NUMBERS

PE: 62702F
PR: 4600
TA: 460014
WU: 460014PT

6. AUTHOR(S)

Paul Mayes, et al

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Rome Laboratory/ERAS
31 Grenier Street
Hanscom AFB, MA 01731-3010

8. PERFORMING ORGANIZATION REPORT NUMBER

RL-TR-94-20
Volume I

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Volume I consists of pages 1 through 276
Volume II consists of pages 277 through 524

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Proceedings of the 1993 Antenna Applications Symposium is a collection of state-of-the-art papers relating to phased array antennas, multibeam antennas, satellite antennas, microstrip antennas, reflector antennas, HF, VHF, UHF and various other antennas.

14. SUBJECT TERMS

Antennas
Satellite Antennas
Broadband Antennas
Microstrip
Reflector
HF, VHF, UHF
Multibeam Antennas
Array Antennas

15. NUMBER OF PAGES

282

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

CONTENTS

WEDNESDAY, SEPTEMBER 22, 1993

DUAL-USE TECHNOLOGY

1. * Keynote: "Dual-Use Antenna Technology - Past and Future," by J. K. Schindler

2. * "Integrated Circuit Active Phased Array Antennas for Millimeter Wave Communications Applications," by G. E. Miller

3. "Applications of Waveguide Arrays in Commercial and Military Radars," by I. Karlsson 1

4. "The Airlink® High Gain Antenna System," by P. Westfeldt, Jr, and J. J. Konrad 23

5. "Focused Aperture Antennas," by C. E. Baum 40

6. "The Radiation Pattern of Reflector Impulse Radiating Antennas: Early-Time Response," by E.G. Farr and C. E. Baum 62

* NOT INCLUDED IN THIS VOLUME

PHASED ARRAYS

7. "The Efficiency of Orthogonal Transform Domain Adaptive Processors for Wideband Multichannel Arrays," by J. S. Goldstein, M. A. Ingram, P. D. Anderson and A. D. Forrest 91

8. "A Picosecond Pulse Measurement System for Nondestructive Evaluation (NDE) Using Step-Frequency Radar," by William H. Weedon, Weng Cho Chew 115

9. "Monte Carlo Techniques for Phased Array Analysis," by D. A. Paschen and S. C. Olson 127

10. "Digital Beamforming," by W. F. Brandow IV, Lt, USAF, W. R. Humbert 139

11. "High Temperature Superconducting Antenna Array Development," by L. D. Poles, J. P. Kenney, E. Martin and J. S. Herd 145

12. "Development of a Dual Frequency Array for ACTS Small Terminal," by M. L. Zimmerman and R. Q. Lee 153

13. "Scanning Characteristics of Stripline-Fed Tapered Slot Antennas on Dielectric Substrates," by D. H. Schaubert 171

14. "A New Waveguide-to-Microstrip Transition," by N. Herscovici 189

THURSDAY, SEPTEMBER 23, 1993

ARRAYS AND REFLECTORS

15. "Advances in Neural Beamforming," by Maj J. Simmers, H. Southall and T. O'Donnell 203

16. "High Performance Microwave Neural Net-Antenna Arrays for Radar Signal Clustering," by A. D. Macedo Filho and H. D. Griffiths 217

17. "A Neural Network Algorithm for Improving the Performance of Distorted Reflector Antennas," by R. J. Bastian and W. T. Smith 227

18. "Theoretical and Experimental Studies of Microstrip Reflect-arrays Used for Mobile and Satellite Communications," by J. Litva, Y. Zhuang and C. Wu 243

19. "A Study of the Synthesis of a Large Communications Aperture Using Small Antennas," by T. Cwik, V. Jamnejad and G. Resch 255

20. "Determining Noise Temperatures in Beamwaveguide Systems," by W. Imbriale, W. Veruttipong, T. Otoshi and M. Franco 277

21. "Feed and Subreflector Displacement Analysis of Shaped Elliptical Dual Reflector Antennas," by M. Im, Y. Chang, M. Sarcione and F. Beltran 301

22. "The New Monopulse Feed Optimized With Antenna Performance," by C. Ming 330

ELEMENTS

23. "Vivaldi Antenna Elements for Microwave Neural Networks," by A. D. Macedo Filho, M. M. Mosso and A. Podcameni 338

24. "Linearly Tapered Slot Antennas and Feed Networks," by R. Q. Lee and R. N. Simons 349

25. "Land Mobile Satellite Vehicular Antenna Development at JPL," by A. Densmore, V. Jamnejad, A. Tulintseff, J. Huang, R. Crist, K. Lee, L. Sukamto and W. Chew 367

26. "Direction Finding With an Electromagnetically Complete Field Sensor," by J. F. Bull 392

27. "The Monopole as a Wideband Array Antenna Element," by D. Collier and H. Shnitkin 406

28. "Slot-Fed Microstrip Antennas With a Microstrip Line," by Y. L. Chen 420

29. "Line-of-Sight of a Helicopter Mounted Antenna," by M. Courtright 431

30. "Measurement Speed and Accuracy in Switched Signal Measurements," by J. Swanstrom and R. Shoulders 450

FRIDAY, SEPTEMBER 24, 1993

ANALYSIS AND MEASUREMENT

31. "Antenna Measurements and Diagnostics in a Unique Bi-Polar Near-Field Facility," by Y. Rahmat-Samii, L. I. Williams and R. G. Yaccarino 455

32. "Productivity Gains Using Multiple-Channel, Multiple-Frequency Measurement Techniques for Testing the E2C Antenna," by J. Scherer, R. Magatagan, J. Miller, G. Amaral and J. Swanstrom 482

33. "Adaptive Antenna Space-Time Processing Techniques to Suppress Platform Scattered Clutter for Airborne Radar," by E. C. Barile, T. P. Guella and D. Lamensdorf 488

34. "Constrained Optimization Method for Shaped Beam Antenna Pattern Synthesis," by Y. Qi, C. Wu and J. Litva 503

35. "Antenna Technology for Wide Angle Scanning," by A. J. Zaman and R. J. Acosta 516

36. * "Isoimpedance Non-Homogeneous Materials in Antenna Applications," by A. I. Knyaz

* NOT INCLUDED IN THIS VOLUME

APPLICATIONS OF WAVEGUIDE ARRAYS IN COMMERCIAL AND MILITARY RADARS

I KARLSSON, ERICSSON RADAR ELECTRONICS AB
MÖLNDAL, SWEDEN

ABSTRACT

Slotted waveguide arrays have proven to be good candidates for a wide range of radar applications. The inherent properties of being a distribution network as well as radiator result in low loss compact designs.

Ericsson Radar Electronics, (ERE), has had the opportunity to study, design and manufacture slotted waveguide arrays for both commercial and military applications.

The interaction between the military and commercial projects has facilitated the development of advanced design tools and techniques for measurements and manufacturing. The paper briefly presents some of the projects and design aids. A frequency and phase steered array antenna for an artillery locating radar is described in more detail.

1. INTRODUCTION

ERE has been involved in several antenna development programs using slotted waveguide array technology covering the S to Ku frequency bands. Some have been awarded by the European Space Agency (ESA) and some by different defence administrations.

Among the programs funded by the European Space Agency are the Synthetic Aperture Radar (SAR) antenna and scatterometer antennas of the first and second European Remote Sensing Satellite[1,2]. Studies and breadboarding of new advanced concepts such as a dual beam array[3], a dual polarized slot antenna[4] and shaped beam antennas[5] are other ESA funded programs.

Among the more recent developments for military programs are the phase scanned AEW radar antenna[6] and the phase and frequency steered array for an artillery locating radar (ALR). The ALR antenna is described below. Arrays for weather radars[7] and commercial side looking airborne radars, monitoring oil spills along the coast, are other programs that have contributed to the experience of slotted waveguides.

It is the combination of both commercial and military programs that has justified investments in design software and test facilities. Some of these programs and investments are briefly presented in the paper.

2. CHARACTERISTICS OF SLOTTED WAVEGUIDES

Slotted waveguide arrays often come out as the winner in trade offs between different antenna techniques. The antennas mentioned above all incorporate broad wall slotted waveguide arrays.

ERE prefers to work with broad wall slotted waveguides rather than edge slot waveguides as the design software is accurate and well verified.

For scanning antennas ridge loaded waveguides are used to avoid grating lobes.

Other attractive characteristics of broad wall slot waveguide arrays are:

- low cross polarisation
- slots in planar ground plane
- low depth (waveguide height)
- low mass
- self supporting design
- low manufacturing cost
- low loss

Accurate design software and manufacturing techniques that allow tight tolerances are absolutely necessary in order to achieve low sidelobes.

ERE uses basically two manufacturing techniques. One is to machine the waveguides and slots in a block of aluminium. The other method is to use extruded waveguides and machine the slots in the waveguides. Both methods have proved to give very good tolerances and high yield. Dipbrazing is used to solder the different antenna parts together.

ERE also has experience of using metallized carbon fibre for manufacturing arrays. The requirements on mass and temperature stability have to be very important (e g space application) to justify the cost of carbon fibre arrays.

3. A PLANAR NEAR-FIELD SCANNER

A planar near-field test facility was required to measure the performance of the 10 meter by 1 meter SAR antenna, which at that time was the largest array ERE had designed. The antenna comprises 5 mechanical panels with resonant slotted waveguides. The electrical design and testing were performed at ERE.

The antenna is mainly built of metallised carbon fibre waveguides and components to achieve low mass and temperature stable performance in orbit.

Dornier GmbH in Germany manufactured the antenna. The test facility has an effective scan area of more than 5 meter by 12 meter. The probe position is controlled with an RMS error of less than 0.1 mm. The frequency ranges from 4 to 40 GHz[8]. Figure 1 shows the test range with the ERS-1 antenna.

The design and test software developed during the program has been further developed in the design of the AEW antenna and the ALR antenna. It would have been very hard to develope and produce the high performance antennas required today without the availability of a high quality near-field test range.

Figure 1 Planar near-field test range with ERS-1 antenna.

4. ARRAY ANTENNA DIAGNOSIS AND CALIBRATION

The need of an accurate array antenna diagnosis tool initiated a method for obtaining the individual element excitations of an array antenna from measured radiation patterns. Access to individual element (slot) excitations makes it possible to compare measured data with design values. Four major applications are identified:

- Design tool. Improve the nominal amplitude and phase distribution during the design phase.
- Antenna calibration. Antennas with phase shifters, variable power dividers and active modules can easily be tuned to compensate for mechanical tolerances and different path lengths in the antenna.
- Element failure diagnosis. Errors due to component failures and manufacturing problems may be located.
- Pattern extrapolation. Test range limitations, with respect to angular coverage, can be overcome by computing patterns at angles where measured data is not available.

Often the measured far-field data is not sufficient to allow an accurate direct transformation back to the element excitations. The measured far-field data is restricted to visible space which does not always contain the entire Fourier domain. A typical example is phased array antennas designed for large scan angels. A similar problem arises during near-field testing of planar antennas in which case a significant far-field domain is restricted by the scanning limitations of the near-field test facility. An iterative procedure has been developed which converges towards the correct solution[9].

The software has been found to be very powerful and it is regularly used to assess antenna designs and calibrate antennas.

5. SLOT CHARACTERIZATION

In the design of slotted waveguide antennas the waveguide is most often modeled as a transmission line where the dominant TE10 mode is propagating. The longitudinal slots cut in the broad wall of the waveguides are normally modeled by normalized shunt admittances. Sometimes a shunt admittance is not sufficient to describe the slot in which case an s-matrix approach can be used. The higher order mode coupling in the TE20 mode can also be included in the waveguide design.

An important part of the design process is to find a model of the slot admittance data which is suitable for computer calculations. The active slot admittance comprises the isolated slot self admittance and a coupling term. The self admittance is a function of slot length, slot offset and waveguide cross section dimensions. The coupling term is dependent on the slot environment and the slot excitation. To find a model of the slot self admittance, that is accurate enough, measurements of slots with different lengths and offsets are required. This is particularly true for ridged waveguides where theoretical modelling is rather complicated[10]. As a result of the many slotted array programs with different waveguide geometry and materials, ERE has developed fixtures and methods that allow very accurate measurements of a large number of slot geometries in a short time. Particularly the models for weakly excited slots have been improved[11]. Curve fitting techniques are used to derive universal polynominals for characterizing the slot in the computer model[12].

Figure 2 shows one slot charaterization fixture.

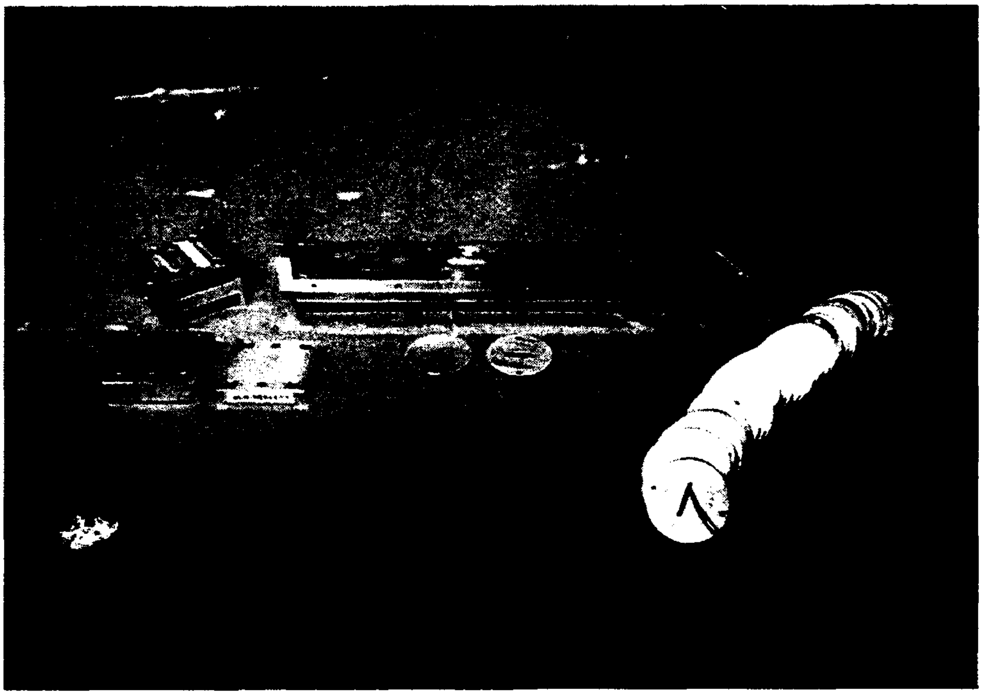


Figure 2 Isolated slot characterization fixture.

6. AN ANTENNA SYSTEM FOR AN ARTILLERY LOCATING RADAR

6.1 Background

The previously described design tools have been used in the design of a phase and frequency steered phased array antenna for use in an artillery locating radar system. The antenna is mounted on a turntable and is designed to be carried on the roof of a tracked vehicle.

A functional model comprising both antenna and turntable has been designed, built and tested. The overall antenna design and performance as well as the performance and requirements of the major components are presented.

The development has been very successful and the measured performance of the functional model antenna meets the target values of the requirement specification.

An artillery locating radar (ALR) has three main functions

- Search for targets (gun, mortar or artillery projectiles) along the horizon.
- Track detected targets to calculate their trajectories and thus the launch positions and impact points.
- Register own artillery fire by tracking projectiles in the last part of their trajectories.

The ALR represents an application where a phased array antenna is almost mandatory to enable the radar to fulfil its main functions. This is because the specific requirements on the radar function such as:

- Very small radar target cross sections, 1000 times below typical aircraft cross sections.
- Very rapid target motion with high vertical velocity.
- Simultaneous tracking of many targets.
- Track while scan mode.
- Very high target data accuracy.

This requires a large antenna with accurate and rapid beam movements in both azimuth and elevation. As indicated in Figure 3 the electrical beam scan sector is wide (>90°) in azimuth but can be limited to a 10° wide sector just above the horizon in elevation.

In the antenna design this is reflected in the choice of frequency scanning in elevation and phase scanning in azimuth. Compared to phase steering in two dimensions this concept is much less costly. Figure 3 demonstrates the ALR concept.

The antenna is mounted on a turntable on the roof of a highly mobile tracked vehicle. The optimum search sector of the antenna for each operating location is adjusted by mechanically tilting and turning the antenna. In transport the antenna is positioned flat towards the roof.

In addition to the stringent requirements on rapid scanning, sidelobes, gain and pointing accuracy the requirements on low cost and low mass also put constraints on the design. The low cost requirement implies conventional production processes and the repeatability of the processes has to be verified. The low mass requirement, generated by the antenna location on the roof of the vehicle, leads to several difficult trade offs as the antenna shall withstand severe environmental conditions during transport and operation.

The design of the ALR antenna has thus been a very challenging task.

The antenna design is presented and the performance of the major components is discussed. The RF signal will be followed on its path through the antenna starting at the power divider input port.

Finally some measured antenna patterns will be showed.

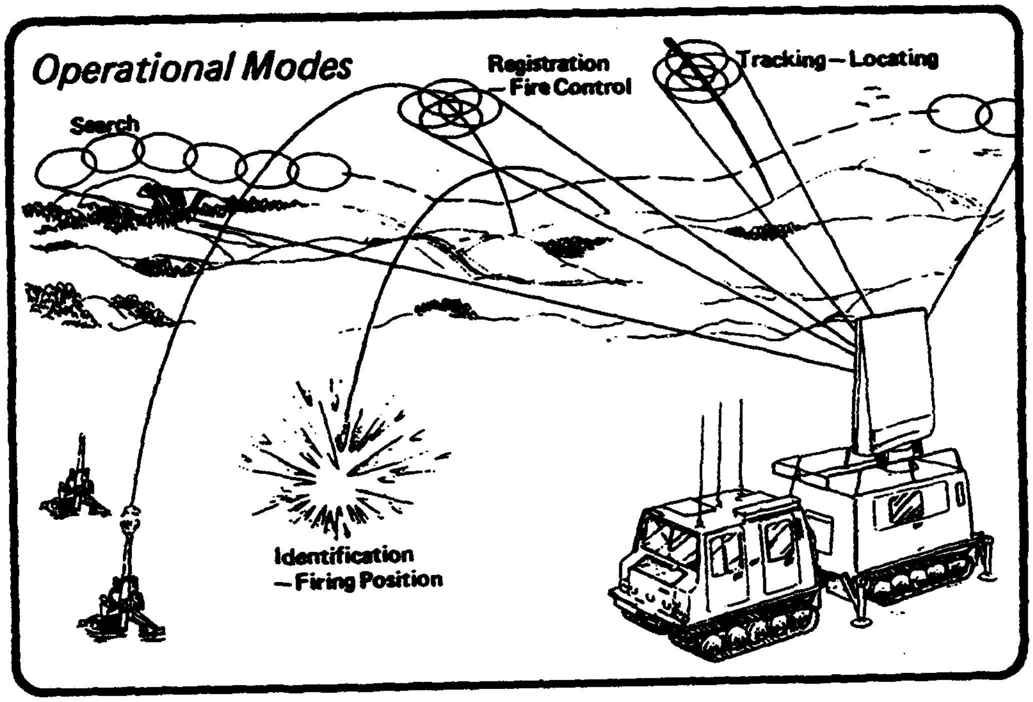

Operational Modes
Registration
—Fire Control
Tracking—Locating
Search
Identification
—Firing Position

Figure 3 Artillery locating radar functions.

6.2 Antenna principal layout

Figure 4 shows the principal layout of the antenna.

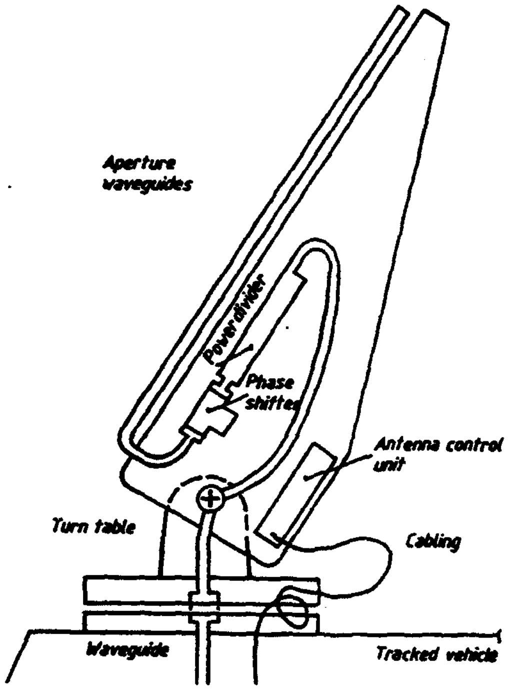

Aperture
waveguides
Powerdivider
Phase
shifte
Antenna control
unit
Turn table
Cabling
Waveguide
Tracked vehicle

Figure 4 Antenna principal layout.

6.3 Power divider

The power divider shall provide a well controlled tapered amplitude distribution across the 48 output ports with an RMS amplitude error less than 0.2 dB. This requirement is set by the azimuth sidelobe constraints. The insertion phase from the input port to each output port may vary considerably. The phase difference is compensated for by the phase shifter settings.

Several power divider concepts were traded off and a corporate feed network was chosen. The design utilizes branch line couplers and the complete network was machined out of two aluminium blocks. Most wall thicknesses were reduced to less than two millimeters to minimize weight. Figure 5 shows the two halves of the network before assembly. The power divider measures 1350 mm x 650 mm x 50 mm and weighs 30 kg. Each half of the network weighs 96 kg before the machining starts. The machining tolerances are better than 0.03 mm on the critical dimensions of the couplers.

The measured amplitude distribution is compared to the requirement specification in Figure 6. The measured average loss of the network is 0.15 dB. This state of the art network was designed and manufactured by ERA Technology Ltd, England.

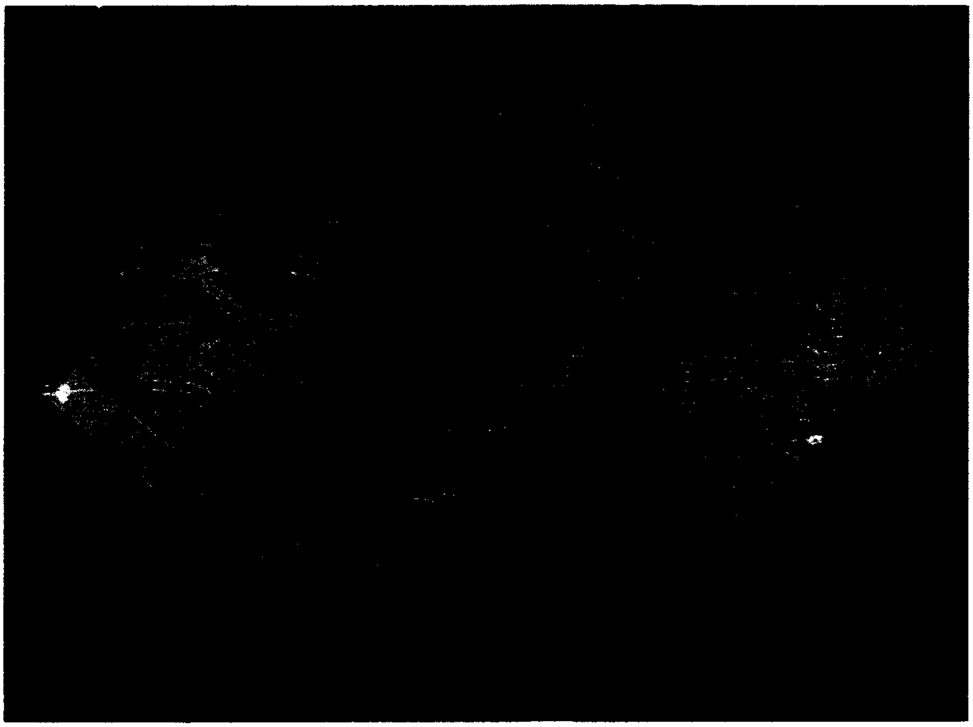


Figure 5 Photo of power divider before assembly.

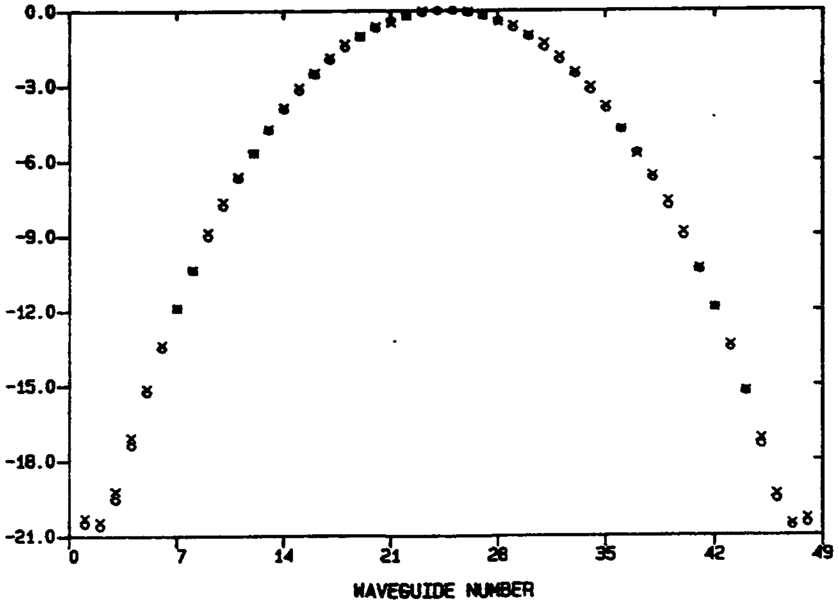


Figure 6 Specified and measured amplitude distributions

o ideal distribution

x measured distribution

6.4 Phase shifter

The phase shifter is a latching, non-reciprocal ferrite device. The magnetization is changed by application of a current pulse through the magnetizing wires. A change in magnetization produces a change in insertion phase with no holding current required. The phase shifter is an 8 bit phase shifter and it is capable of differential phase shift from 0 to 360° in approximately 1.4° increments.

The phase shifter has a phase accuracy of 3° RMS. The repeatability is better than 0.75°. The insertion loss at room temperature is better than 0.5 dB.

A special feature of the phase shifters is that the driver circuitry of each phase shifter accepts a common message from the control unit. The method of distributed beam steering reduces the size of the central control unit and requires less cabling.

Each phase shifter has an E^2 prom for storing phase corrections to compensate for different path lengths in the power divider and other phase errors. The phase corrections are readily attained from the diagnosis and calibration software during the final test.

6.5 Aperture waveguides

From the phase shifters the signal enters into the aperture waveguides. The aperture comprises 48, 2.3 m long ridged waveguides with 64 slots in each waveguide. The waveguides are non resonant and terminated with matched loads. Broadwall slots have been chosen, partly due to the availability of very accurate inhouse software for synthesis and analysis of waveguide arrays with broad wall slots.

The requirements on large scan angles (±45°) in azimuth dictated the use of ridged waveguides. Standard rectangular waveguides would allow only about 20° scanning without generating grating lobes.

The choice of waveguide dimensions was also dictated by the required frequency scanning in elevation which is directly related to the guide wavelength. To achieve the scan range the waveguide had to operate relatively close to the cutoff frequency.

The guide wavelength of a ridged waveguide close to cutoff is very sensitive to the waveguide profile. The tolerances on the most critical dimensions were set to 0.03 mm to maintain a good elevation pointing accuracy.

The slot positions relative to the centerline of the waveguide determines the excitation. For a ridged waveguide the field is concentrated at the ridge and even small errors in the slot offset can result in serious excitation errors and high sidelobes. The worst type of errors are systematic errors. For example, if the reference plane has an offset error when the slots are machined, every other slot will couple to much or to little power respectively. This will result in grating lobes.

A systematic slot offset error of 0.04 mm would result in a 20 dB grating lobe in the elevation pattern.

The problem of maintaining tight tolerances was one of the major concerns in the beginning of the program. Several ways of producing and machining the waveguides were investigated. The waveguides for the present antenna model were extruded and machined by MIFA Aluminium bv, Holland.

Another concern was the characterization of slots that are offset less than half a slotwidth from the centerline. This is the case for several slots near the waveguide input. This problem was overcome by very accurate measurements of weakly excited slots[13].

The electrical design was done in an iterative process. The slot excitation and the amount of power coupled into the end load were inputs to the synthesis which was done at center frequency and included mutual coupling effects.

The derived slot positions were in the next step used for analyses over the frequency band of interest. As the array comprises over 3000 slots and the mutual coupling effects were considered each frequency to be analysed required long computer time. The optimum antenna performance was achieved when only 4 % of the input power was coupled to the end load.

The design was also supported by measurements on individual waveguides to verify the calculations. Figure 7 shows a picture of the antenna aperture.

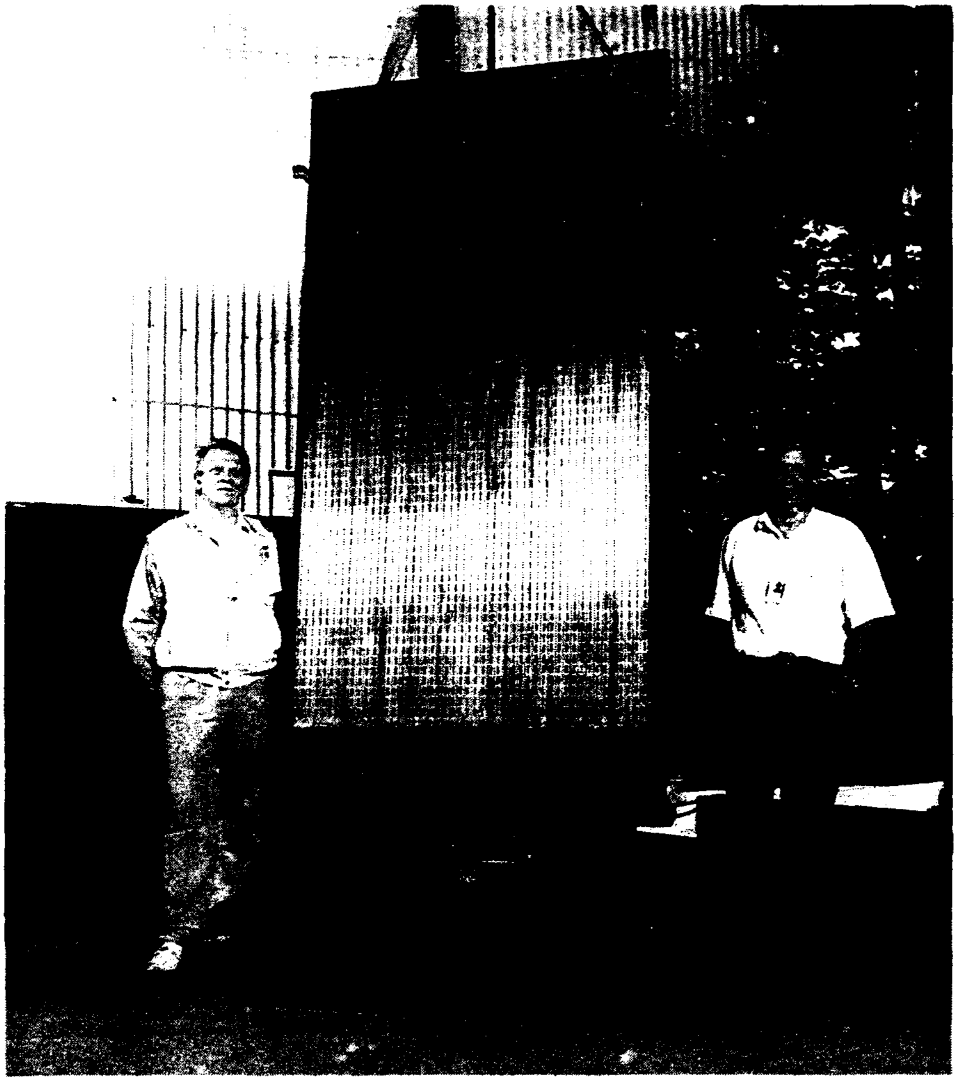


Figure 7   Antenna mounted on the turntable.

6.6 Antenna control unit

The antenna control unit is a communication interface between the signal and data processing unit and the phase shifters. It also takes care of the built in test signals and signals from different sensors in the antenna.

6.7 Antenna pointing

To enable the radar to accurately determine the position of the launching artillery or impact point of the detected target the position of the antenna has to be known very accurately. The antenna beam pointing has also to be known within 2 mrad. A sensing system is implemented to track the antenna movements to achieve this under all possible environmental conditions and to avoid a heavy antenna and turntable structure.

6.8 Antenna performance

Figure 8 shows a measured 3D plot of the antenna pattern for a beam scanned 30° in azimuth. The concentration of sidelobe peaks in the principal planes, are typical for a rectangular antenna. In all other areas the sidelobe levels are below the noise level. Figure 9 shows a measured antenna pattern for an elevation cut at the center frequency. Figure 10 shows a measured pattern for an azimuth cut, at 30° scan angle.

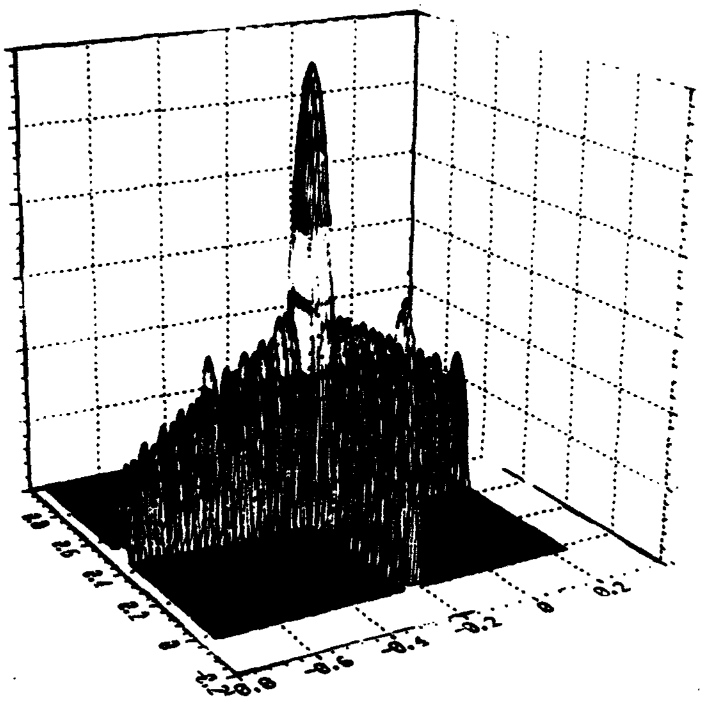


Figure 8 3D plot of measured antenna pattern.

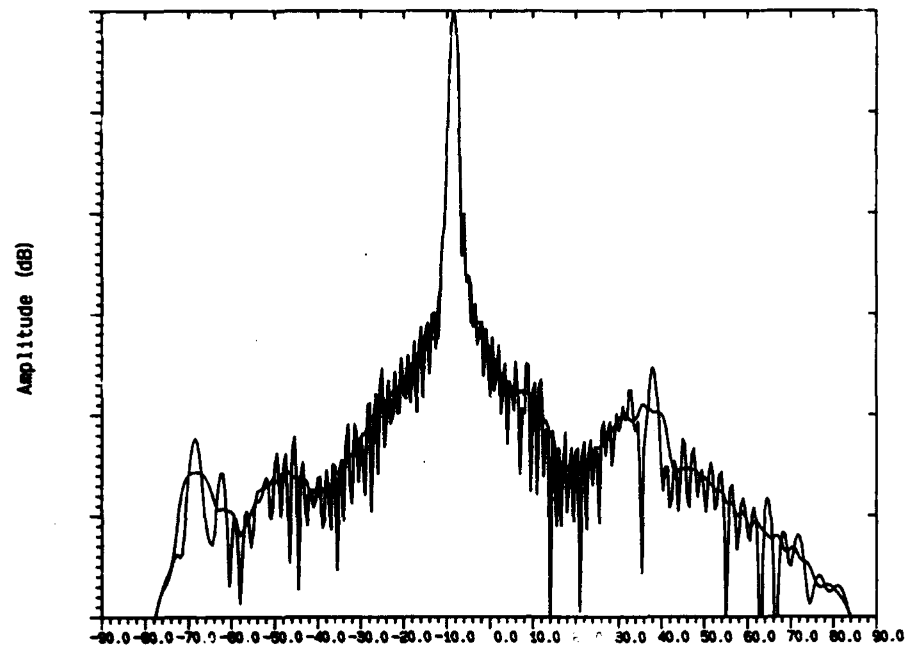


Figure 9 Elevation cut measured data.

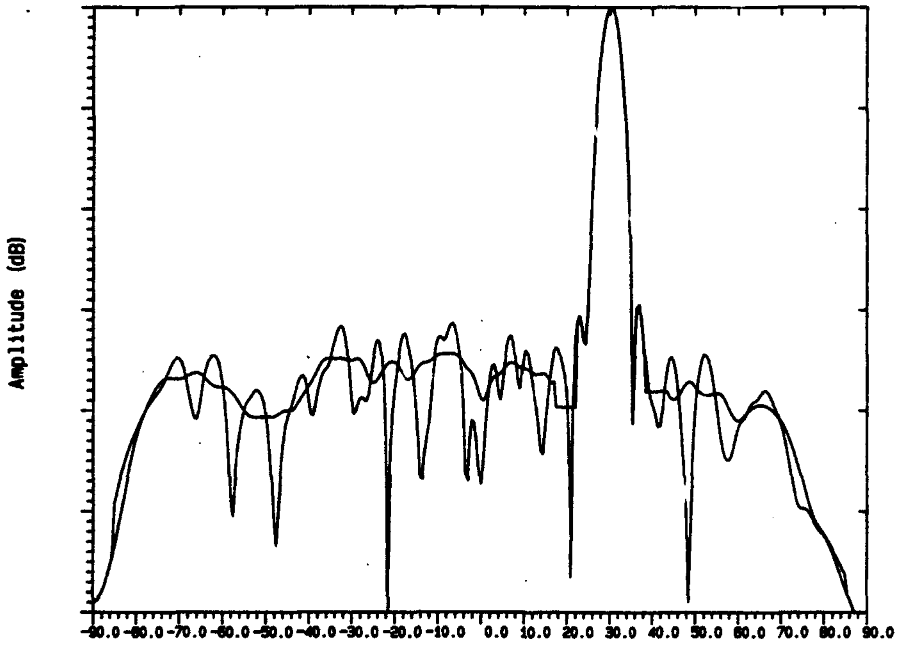


Figure 10 Azimuth cut measured data.

7. SUMMARY

The combination of commercial and military programs has allowed ERE to invest in design software and test facilities which are required in the design of advanced slotted waveguide arrays. The paper presents some of these design aids and describes in more detail the design of an antenna for an artillery locating radar. A similar antenna has been proposed for a weather radar application.

8. REFRENCES

1 Petersson, R, Ingvarson, P, (1988), "The planar array antennas for the European remote sensing satellite ERS-1", Proceedings 18th European Microwave Conference (September 12-16, 1988, Stockholm), pp 289-294.

2 Petersson, R N O, (1992), "Antennas for space scatterometers and SAR's", Conference proceedings MM 92 (October 14-15, 1992, Brighton), pp 249-254.

3 Bonnedal, M, Karlsson, I, K van't Klooster, (1991), "A dual beam slotted waveguide array antenna for SAR applications", Proceedings IEE Seventh International Conference on Antennas and Propagation ICAP'91 (April 15-18, 1992, York, UK), London, pp 559-562.

4 Josefsson, L, (1992), "A dual polarised slotted waveguide array antenna, URSI Int. Symp. Electromagnetic Theory, 17-20 Aug., Sydney,1992, pp 1-3.

5 Rengarajan, S R, Derneryd, A G, (1993), "Application of compound coupling slots in the design of shaped Beam Antenna Patterns", IEEE Transactions on Antennas and Propagation, Vol. 41, No 1, pp 59-65.

6 Lagerlöf, R, Josefsson, L, (1990), "A phased array radar antenna for airborne early warning", Proceedings Military Microwaves '90, (July 11-13, 1990, London), pp 257-261.

7 Josefsson, L, (1991), "Phased array antenna technology for weather radar applications, Proc. 25th Int. Conf. on Radar Meteorology, (June 24-29, Paris, France).

8 Malmborg, P, Molker, A, (1988), "The Ericsson-ESA near-field facility", Proceedings of the 11th ESTEC Antenna Workshop on Antenna Measurements (June 20-22, 1988, Gothenburg, Sweden), pp 243-257.

9 Johansson, M, Svensson, B, (1990), "Array antenna diagnosis and calibration", AMTA Proc., (pp 10:27-10:32, Philadelphia, PA).

10 Falk, K, (1988), "Admittance of a longitudinal slot in a ridge waveguide", Proc. IEE, Pt H, No 4, August 1988, pp 263-268.

11 Rengarajan, S R, Derneryd, A G, (1992), "Improved scattering model for weakly coupled broadwall slots", IEEE Antennas and propagation society international symposium digest, (July 20-24, Chicago, Illiniois, USA, pp 41-44).

12 Derneryd, A G, Lorentzon, T C, (1991), "Synthesis of a phased array with single-ridge waveguides", Proceedings Antenna '91 (August 12-16, 1991, Fårö, Sweden), pp 319-328.

13 Derneryd, A G, Lorentzon, T C, (1991), "Design of a phase/-frequency scanned array antenna with non-resonant slotted ridge waveguide elements", IEEE Antennas and Propagation Society International Symposium Digest (June 24-28, London, Ontario, Canada, 1991), pp 1728-1731.

THE AIRLINK® HIGH GAIN ANTENNA SYSTEM

Patrick Westfeldt, Jr.
John J. Konrad
Ball Aerospace and Communications Group
Broomfield, Colorado 80038

ABSTRACT

Developed first for commercial telecommunications, the AIRLINK® High Gain Antenna System has become the market leader in commercial aircraft installations. Two side-mounted phased arrays are employed on a single aircraft, using conformal passive apertures with active electronics modules mounted inside the skin. Coverage is achieved throughout the INMARSAT coverage region, and the system meets all ARINC, FAA, and INMARSAT requirements. The commercial system has also been reconfigured as a military secure telecommunications link (STU-III). The reconfigured system was demonstrated successfully in May of this year, resulting in the development of a mobile demonstration vehicle and a number of probable system deployments.

The passive conformal aperture is a multilayer printed circuit antenna and feed system less then 0.4 inches in overall thickness. Radiating elements are microstrip patch antennas covering both transmit and receive bands, while the feed is a buried microstrip structure sharing the radiating aperture. RF interfaces from the aperture elements to the electronics modules are accomplished with two custom cable bundles. The system architectures achieves balance between the competing demands of RF performance, reliability, modularity and low production cost to be competitive in commercial markets. Antenna pattern and system performance are presented in this paper.

1.0 INTRODUCTION

As an example of dual use technology, the AIRLINK® system may be unusual in that it was developed first at Ball Aerospace for commercial and general aviation. The development actually began in 1983 with an INMARSAT feasibility study, followed by system and hardware development funded internally and under a second INMARSAT contract. An important milestone was reached in 1991 with a business alliance with Collins Air Transport Division, whereby Collins supplies the avionics and acts as the single source for the complete SATCOM system. At the present time, AIRLINK® has been commissioned on large numbers of commercial and general aviation aircraft, including all versions of the Boeing 7XX series aircraft and Gulfstream II, III, and IV. The backlog of additional installations is significant, following the cycles of new aircraft deliveries and airline retrofit planning.

The antenna system represents mature microstrip phased array technology at Ball Aerospace. The key design issues involved execution of a low-cost producible package, while providing the required performance, reliability and maintainability. Some of these issues are discussed in the following sections. The antenna concepts are not difficult, but the low-cost producible design has allowed Ball to become the leader in this market.

2.0 AIRLINK® SYSTEM DESCRIPTION

AIRLINK® is an airborne satellite communications system fully compliant with multi-organizational standards such as ARINC 741, FAA, RTCA-DO-160C and INMARSAT multi-channel voice/data technical requirements. The system capability enables an aircraft to use a global aeronautical communication network that consists of three components, the airborne earth station (AIRLINK®), geostationary satellites, and ground earth stations. The space segment consists of four INMARSAT satellites distributed in geosynchronous orbits that cover the major ocean regions as shown in Fig. 1. This distribution of satellite positions provides for worldwide SATCOM coverage except in some areas of the polar regions. The ground segment consists of three COMSAT Ground Earth Stations that interconnect with the Public Switched Telephone Network, providing worldwide service area coverage for the INMARSAT satellites. Aircraft/satellite communications are at L-band, while the ground station/satellite communications are at C-band.

Fig. 2. is a diagram of the AIRLINK® antenna system, consisting of two sidemounted apertures conformal to the aircraft skin, and six internally-mounted RF modules with associated cabling and connections for RF, power and control. The passive radiating apertures are connected by RF cable bundles to the Beam Steering Units(BSU), which contain 3-bit phase shifters for each radiating element, as well as an amplitude-tapered power divider/combiner for each aperture. The common port of the BSU is connected by a short cable run to the LNA/Diplexer module where the transmit and receive paths are split. Downstream from the LNA a simple starboard/port combiner leads to the main receive port. The transmit signal from the High Power Amplifier (HPA) is switched between starboard and port signal paths by means of a high power RF switch.

The architecture shown in this diagram is driven by the requirement for minimum drag from the radiating aperture, along with the availability of space inside the aircraft for mounting the RF electronics modules separate from the aperture. The resulting configuration provides a high degree of maintainability and reliability at the lowest possible

cost. Performance is compromised somewhat because of the passive BSU circuitry and cabling between the aperture and LNA/diplexer.

The coverage plan for the two sidemounted apertures is shown in Fig. 3. At a mounting angle of 45o, the aperture coverage overlaps at the zenith in a region (termed hysteresis) where a handoff routine between apertures is implemented. As expected, low elevation coverage on the nose and tail is limited.

3. DUAL USE

In a more recent internal development at Ball Aerospace, the AIRLINK® system has been reconfigured for secure military communications. The system is designed to offer worldwide, clear, reliable, two-way, air-to-ground or ground-to-air secure or non-secure voice and/or data communication. The secure interface unit of the audio/data subsystem performs analog/digital signal conversions between the Secure Terminal Unit (STU-III) and the Satellite Data Units of the SATCOM transceiver. The STU-III is an analog voice signal device used for both clear and secure voice and data communications.

The system is currently installed on a USAF Avionics Test Bed aircraft for operations testing and evaluation. Numerous deployments of the system are pending. Furthermore, the system is available for demonstration on a Ball Aerospace mobile test bed, which can demonstrate secure voice, video, FAX and PC transmissions.

4. APERTURE DESCRIPTION

The multilayer printed circuit antenna aperture is shown in Fig. 4, in which the circuitry on four different layers are superimposed. It can be readily seen that the full available

aperture is not used, resulting in a some compromise in performance. The space is used instead for the RF signal lines from the bundled coaxial inputs to each radiating element. As shown below, this choice enhances the low cost and producibility of the multilayer aperture. Both RF signal lines and polarization hybrids are implemented in buried microstrip, and are unshielded in the radiating aperture. The radiating elements themselves are two-layer stacked microstrip patches, covering the full range of transmit and receive bandwidth (about 10%).

Fig. 5. is a cross-sectional view of the radiating aperture, having a total thickness 0.375 in. The bottom layer contains the RF signal lines and the polarization hybrids, fed in a surface launch from the coaxial connectors in two bundles. Within the polarization hybrid the signal actually jumps from the trace layer on the bottom board to the back side of the lower (driver) element board, so that the feed ribbon can be installed easily prior to any laminations. The upper (driven) element is etched on a separate board, and the entire four-layer lamination (with a thin laminated radome) is accomplished at one time.

The aperture assembly is about as simple as a multilayer microstrip antenna can be. The installation of the two coaxial connector bundles (shown in Fig. 6) is actually the most difficult part of the process. There were certainly design alternatives that would have provided marginally greater performance, but at higher levels of complexity and production cost.

5. PERFORMANCE DATA

Pattern, gain, G/T and ERP data for the AIRLINK® antenna subsystem are shown in the following figures. Fig. 7. is a typical pattern plot with the antenna scanned to 45° along the long (azimuth) dimension of the array. Testing and qualification of AIRLINK® also requires patterns where the antenna is constantly scanning as it rotates, so that the peak of beam points to the same angle. An example of this type of peak gain pattern is shown in Fig. 8. The same type of scanning also produces contour plots from data taken over half-space. Fig. 9 is a plot of the 12.0 dBic gain contour for AIRLINK® S/N 120. Single antenna data is also transformed into aircraft coordinates for particular commissionings to determine ERP and G/T coverage. Fig. 10 and 11 show aircraft coverage contours at 25.5 dBW ERP and -13.0 dB/K° G/T.

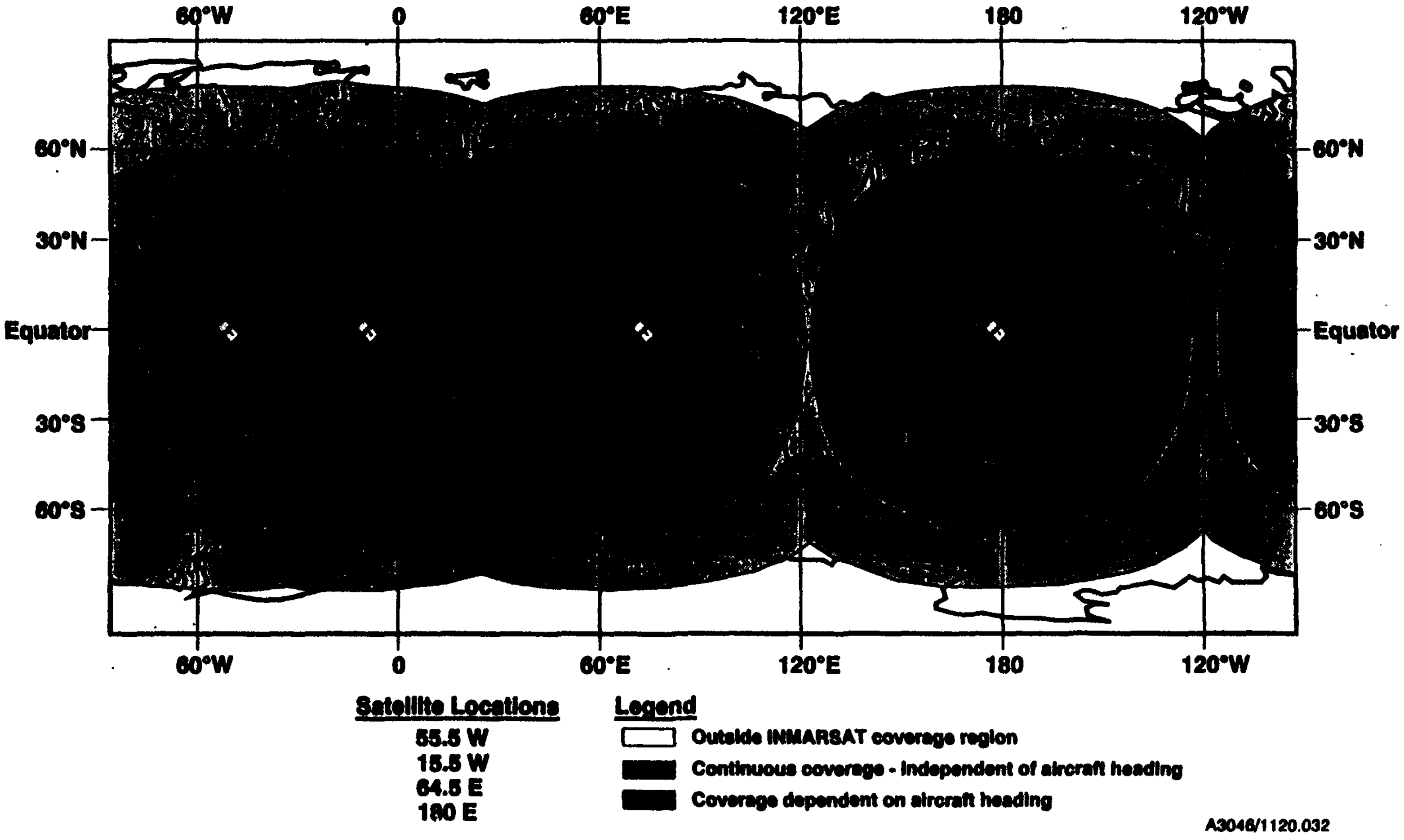

60°W
0
60°E
120°E
180
120°W
60°N
30°N
Equator
30°S
60°S
Satellite Locations
55.5 W
15.5 W
64.5 E
180 E
Legend
Outside INMARSAT coverage region
Continuous coverage - Independent of aircraft heading
Coverage dependent on aircraft heading
A3046/1120.032

Fig. 1. AIRLINK® GLOBAL COVERAGE

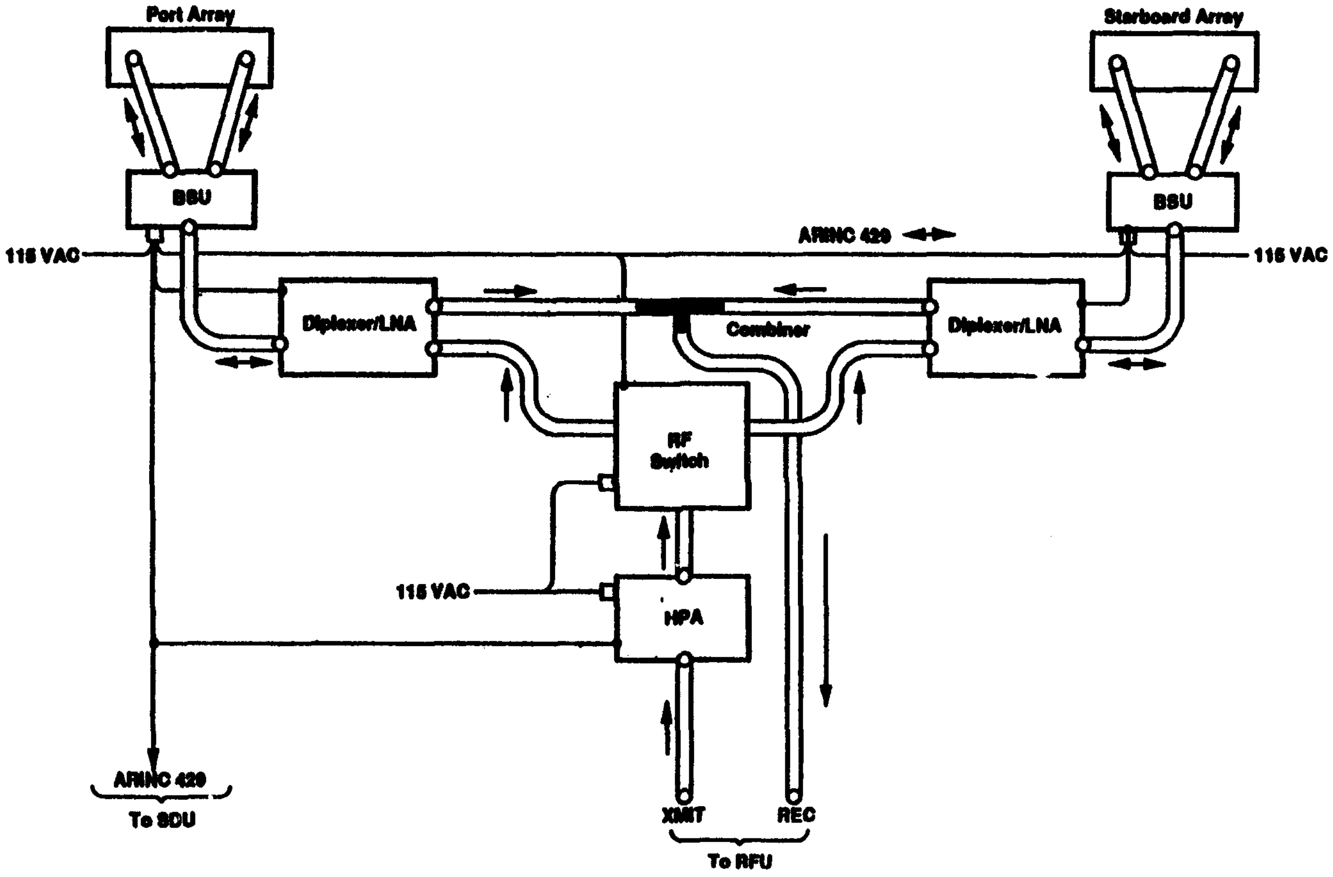
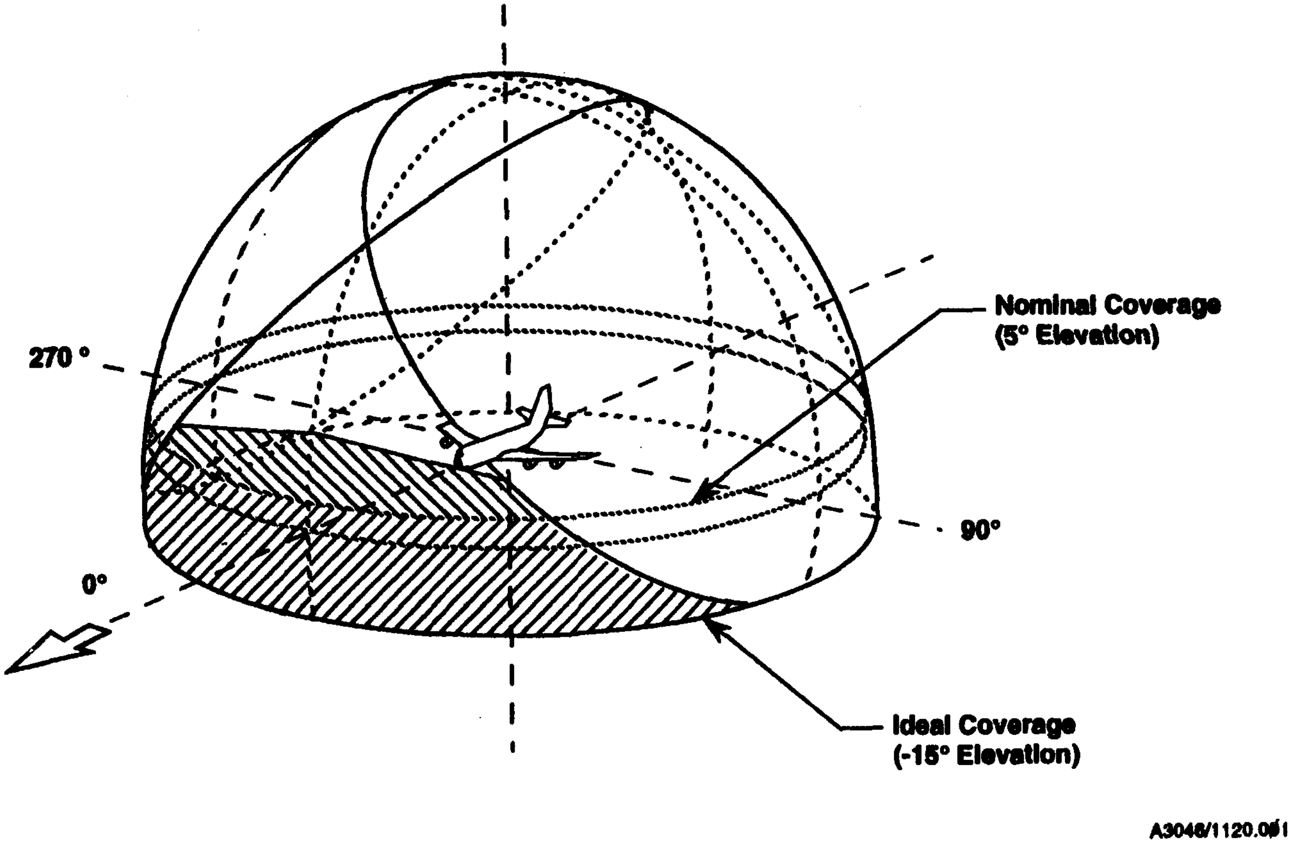

Port Array
Starboard Array
BSU
BSU
115 VAC
115 VAC
ARINC 429
Diplexer/LNA
Diplexer/LNA
Combiner
RF Switch
115 VAC
HPA
ARINC 429
To SDU
XMIT
REC
To RFU

Fig. 2. AIRLINK® ANTENNA SYSTEM CONFIGURATION



270 °
0°
90°
Nominal Coverage
(5° Elevation)
Ideal Coverage
(-15° Elevation)
A3046/1120.001

Fig. 3. AIRLINK® COVERAGE DIAGRAM

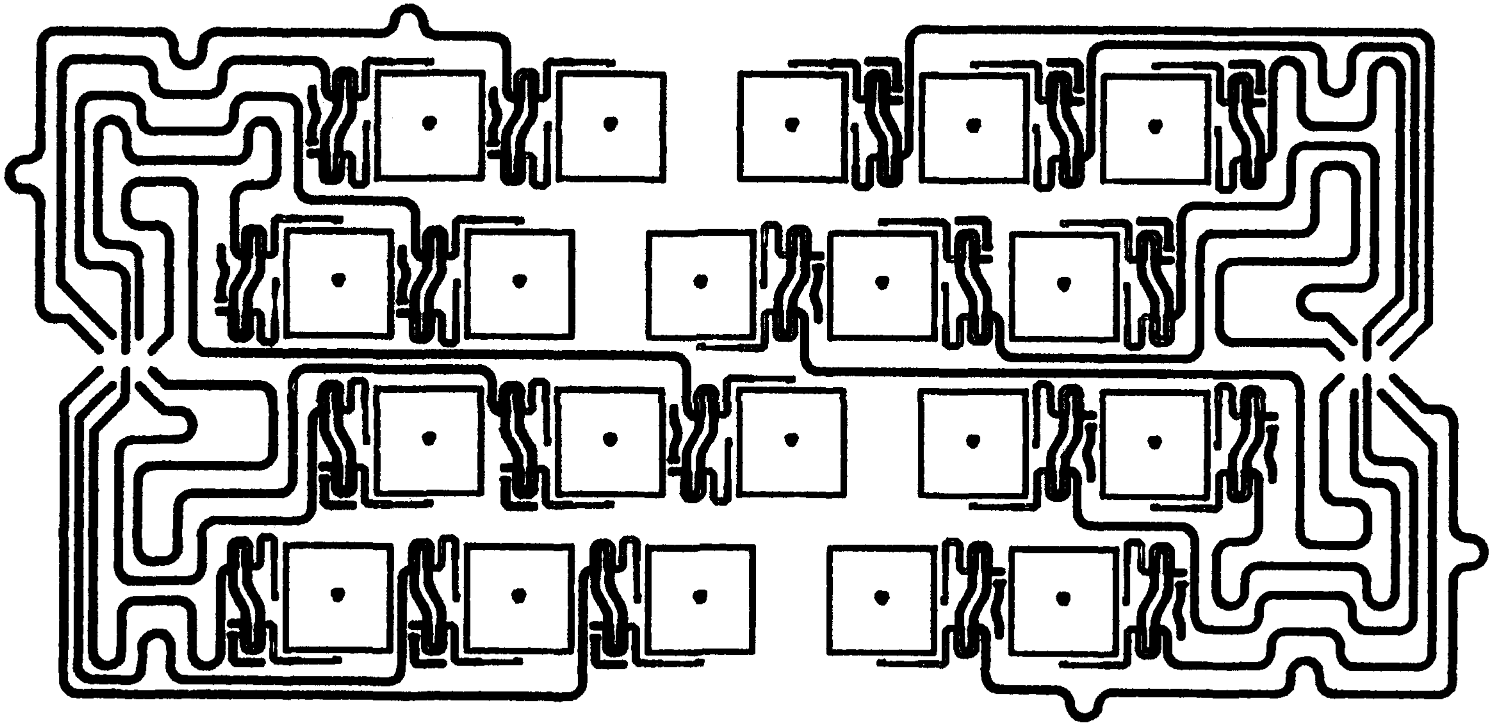


Fig. 4. AIRLINK® APERTURE LAYOUT

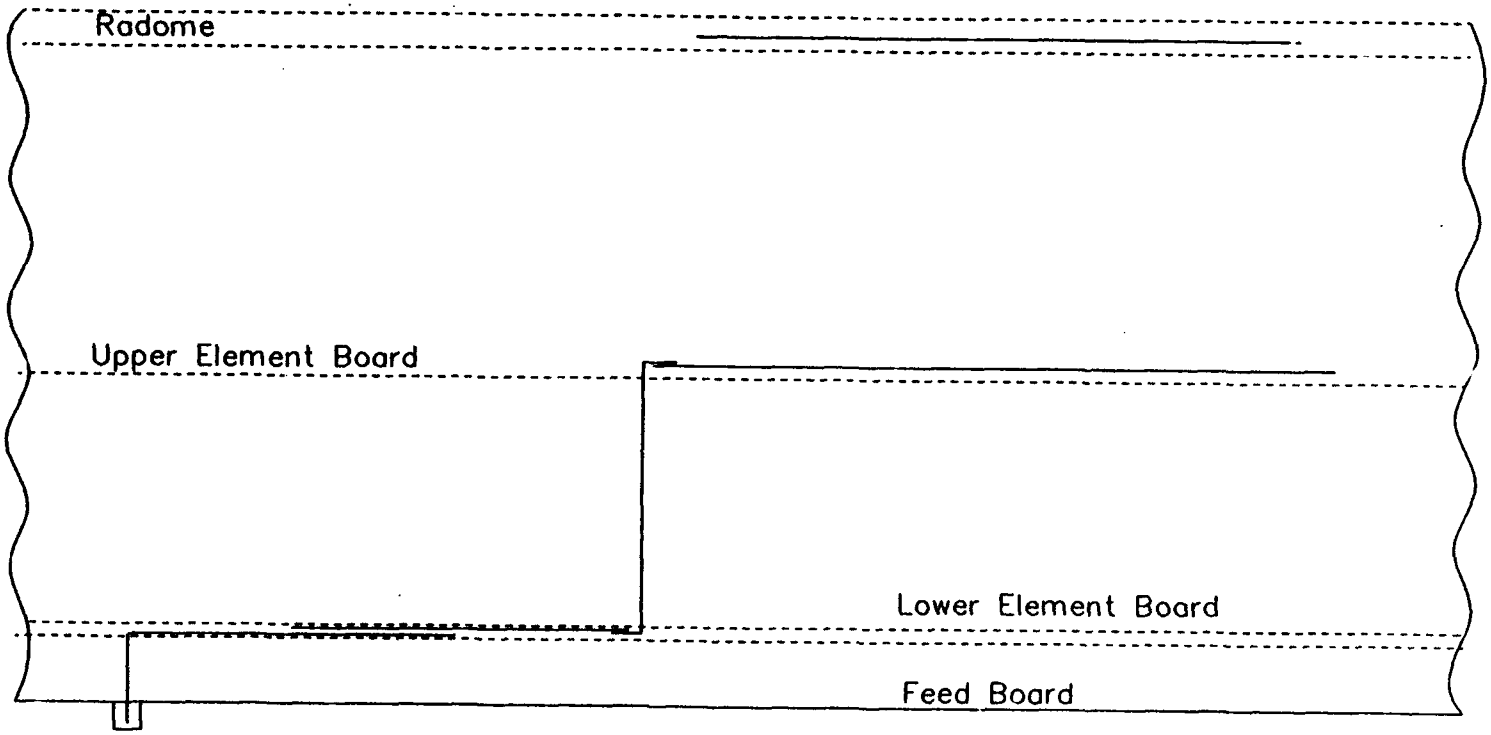

Radome
Upper Element Board
Lower Element Board
Feed Board

Fig. 5. APERTURE CROSS SECTIONAL VIEW

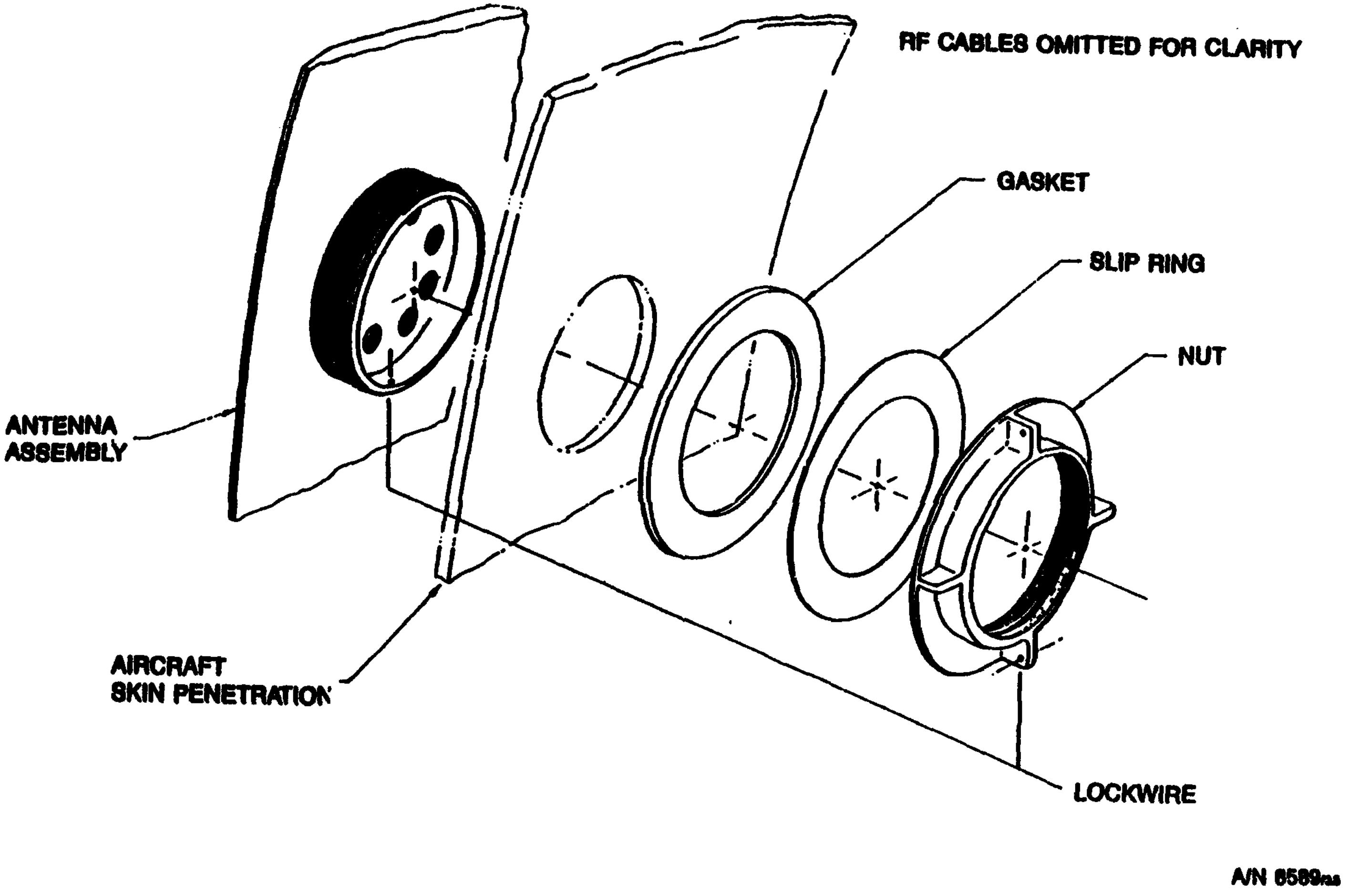

RF CABLES OMITTED FOR CLARITY
GASKET
SLIP RING
NUT
ANTENNA ASSEMBLY
AIRCRAFT SKIN PENETRATION
LOCKWIRE
A/N 6569ns

Fig. 6. RF CABLE CONNECTOR ASSEMBLY

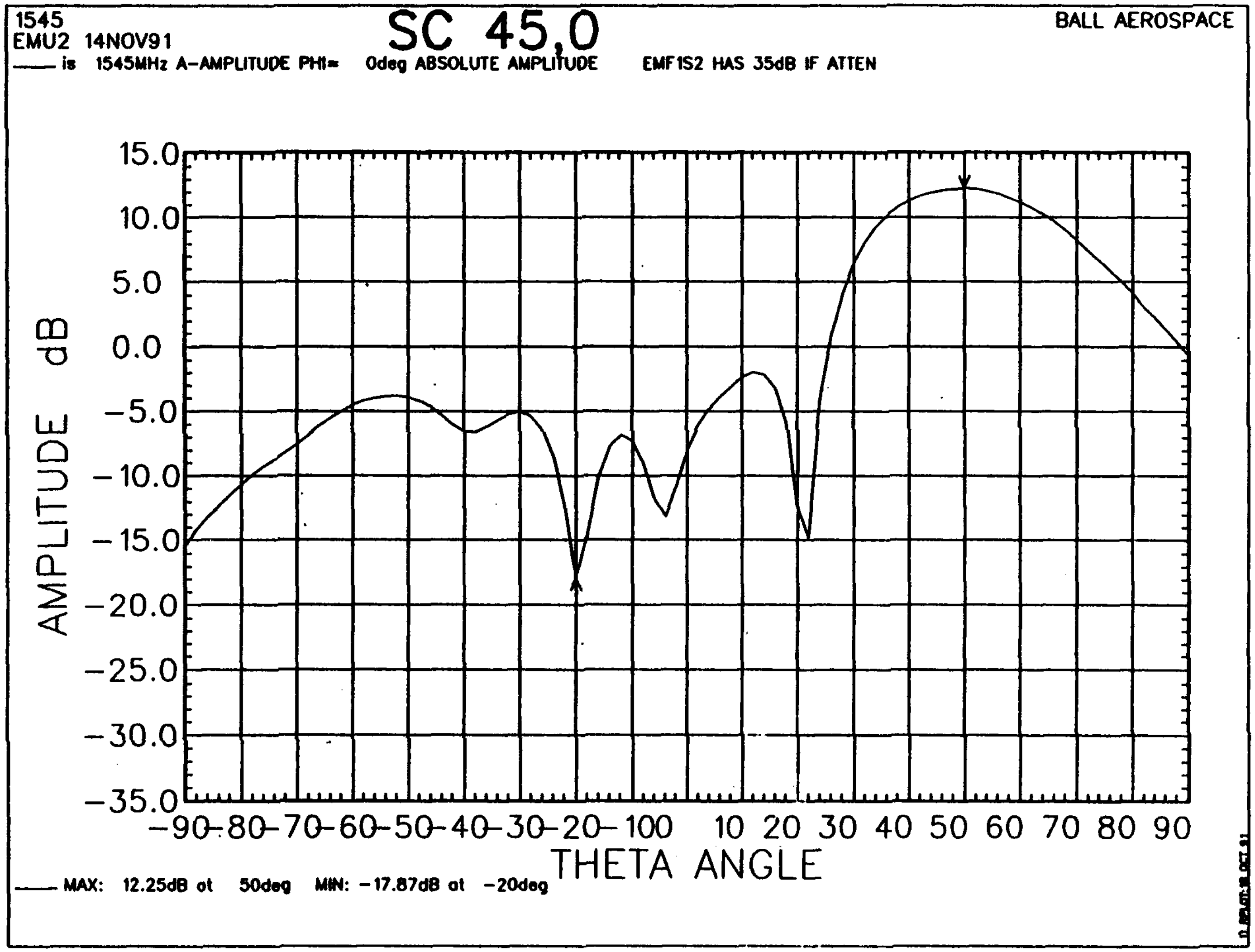


Fig. 7. AIRLINK® PATTERN SCANNED 45°

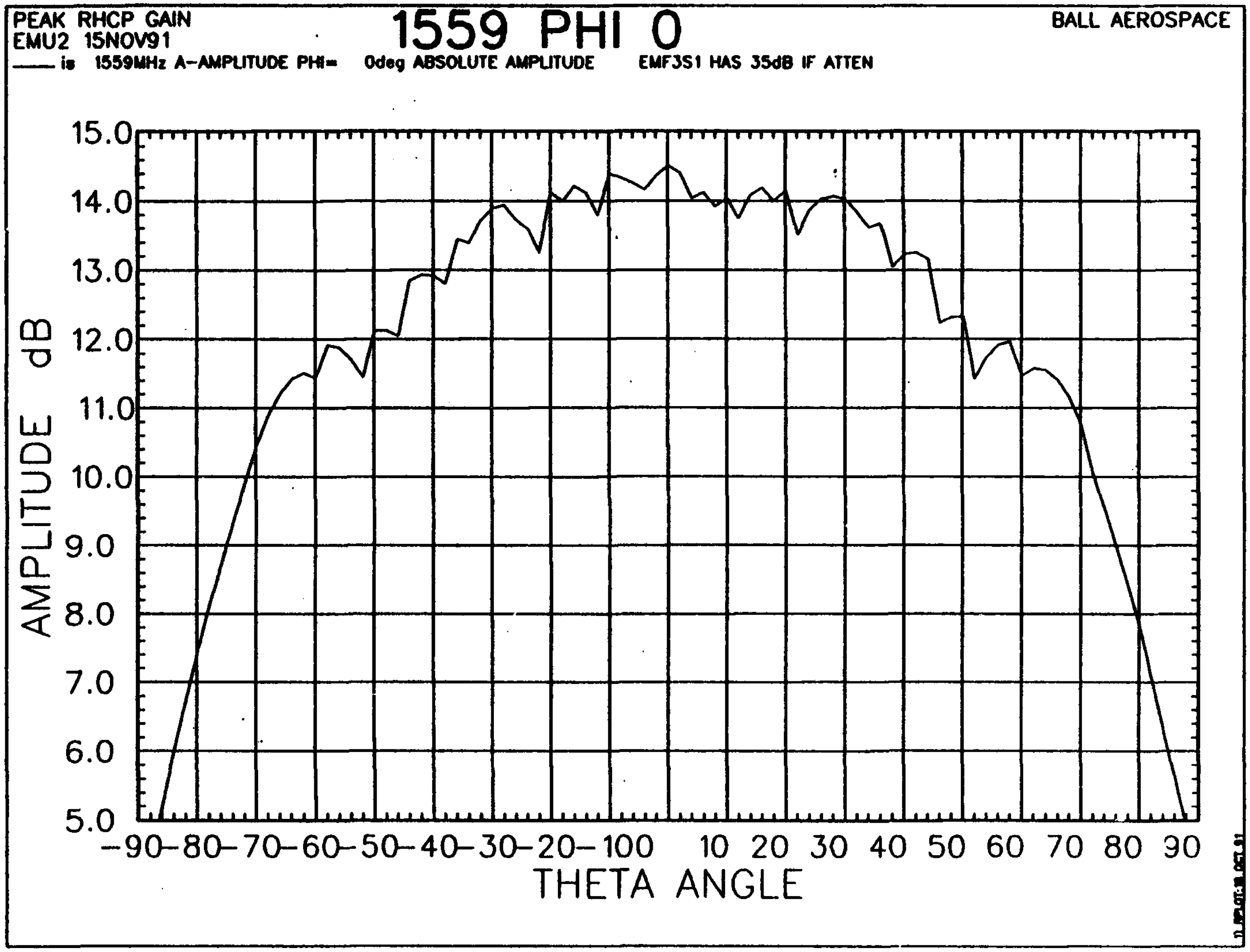


Fig. 8. PEAK GAIN PATTERN, AZ=0°

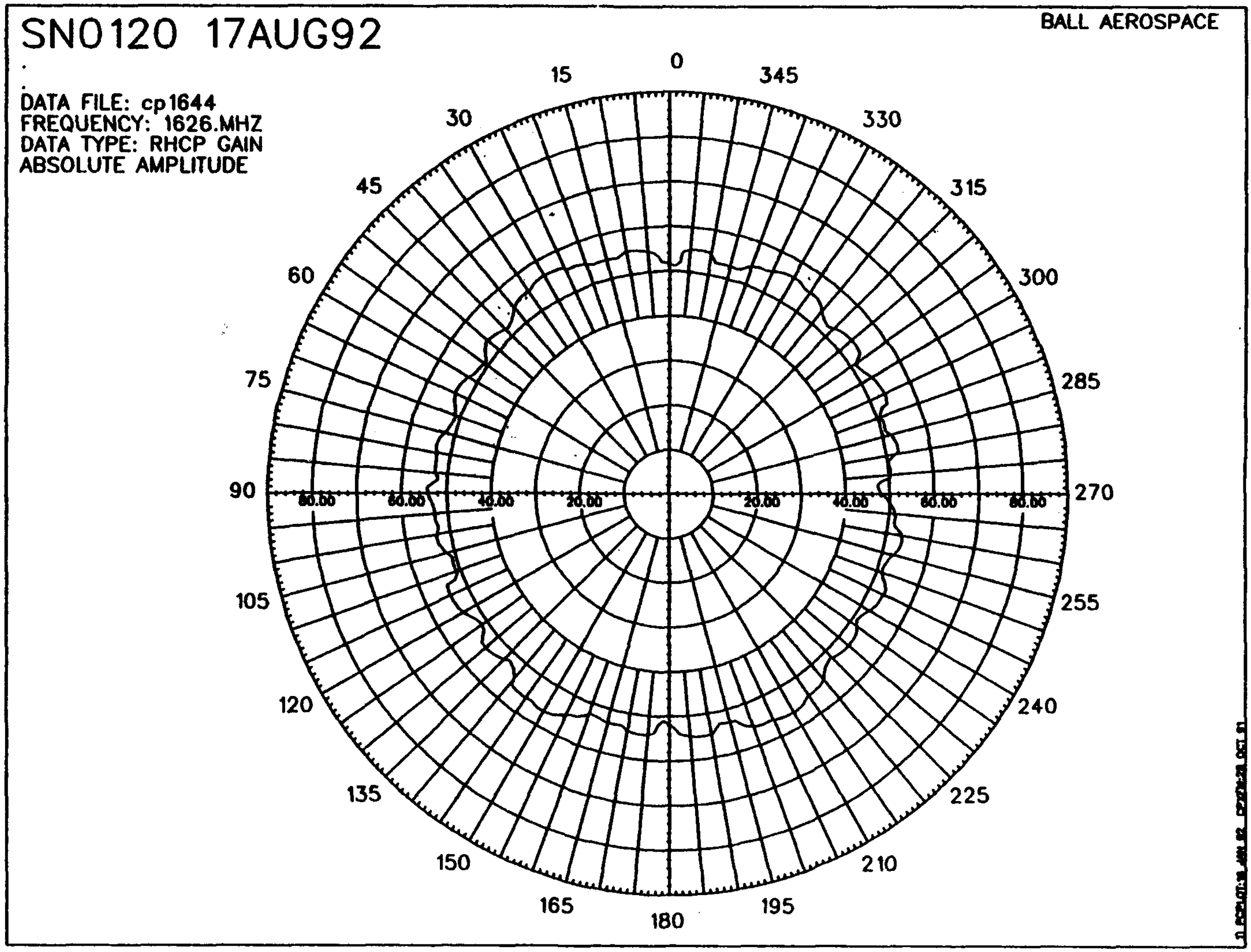


Fig. 9. AIRLINK® 12.0 dBic CONTOUR

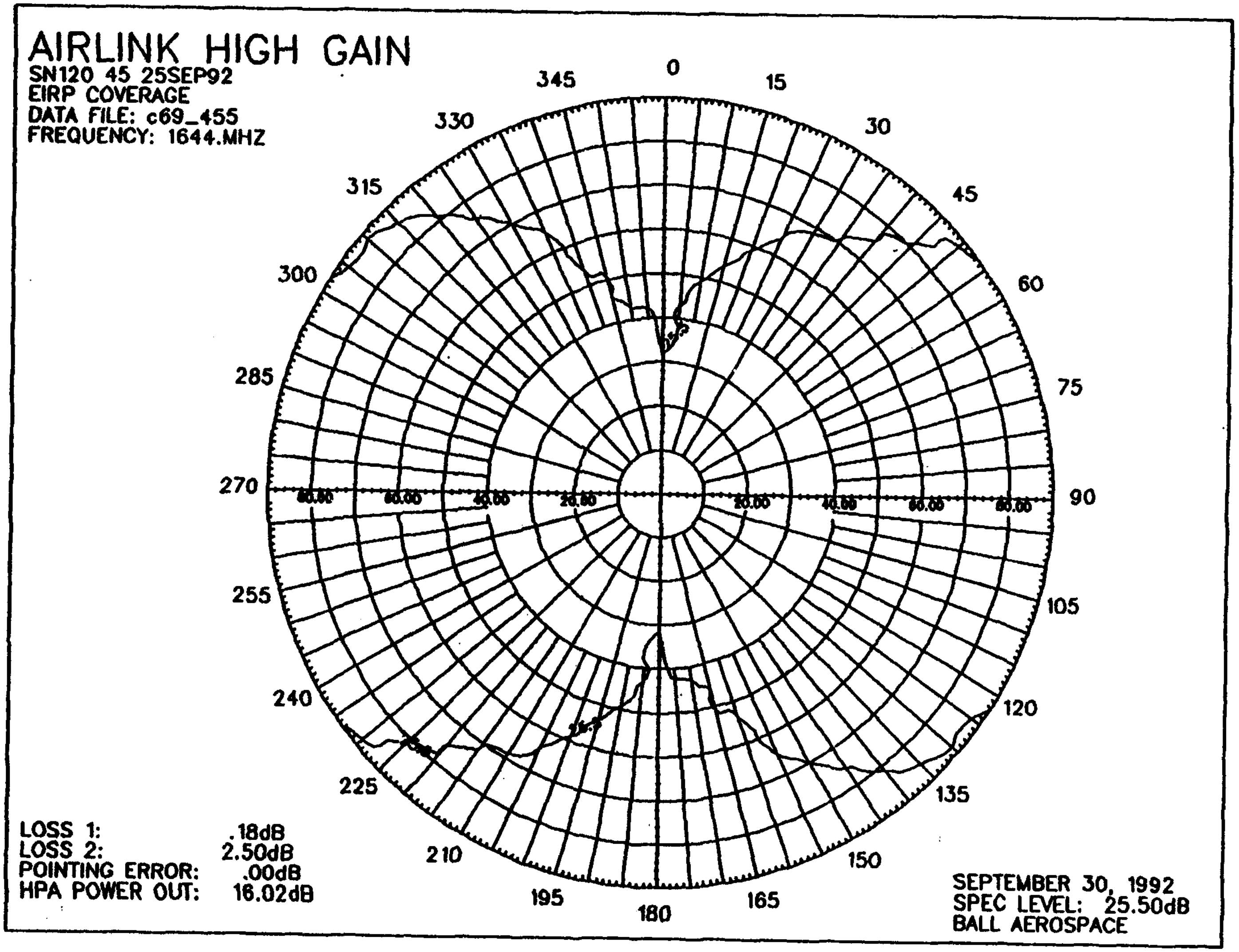


Fig. 10. AIRLINK® 25.5 dBW CONTOUR IN AIRCRAFT COORDINATES

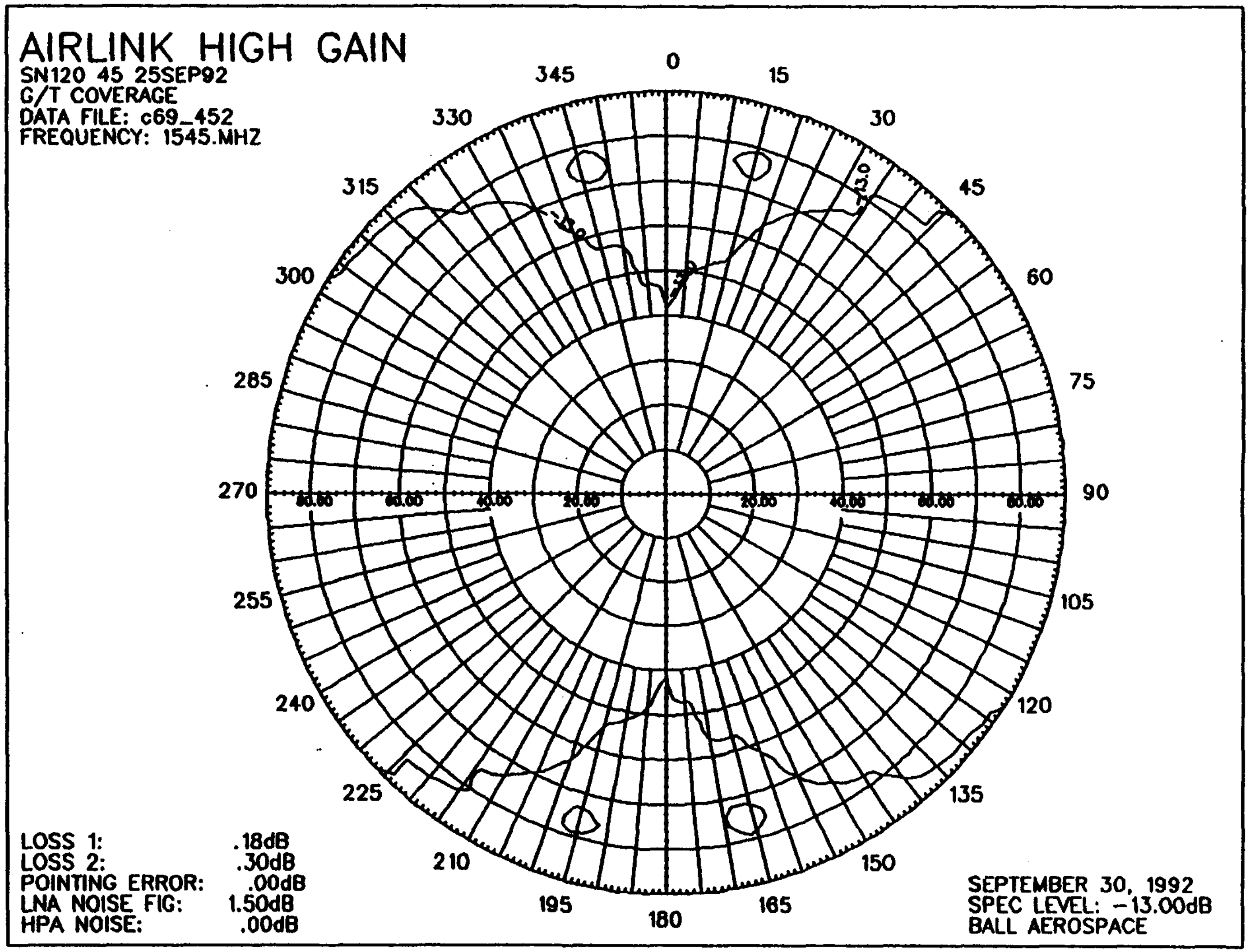


Fig. 11. AIRLINK® -13.0 dB/K° CONTOUR IN AIRCRAFT COORDINATES

FOCUSED APERTURE ANTENNAS*

Carl E. Baum

Phillips Laboratory/WSR
3550 Aberdeen Avenue SE
Kirtland Air Force Base, New Mexico 87117-5776

ABSTRACT

The electromagnetic fields from aperture antennas can be represented as integrals over the aperture electric field. Maximizing the fields at an observer defines a focused aperture. In this case, the integrals simplify and the spatial and frequency parts conveniently separate. This makes the results also conveniently expressible in time domain.

---

* From Sensor and Simulation Note 306, May 1987.

1. INTRODUCTION

A previous paper [2] has considered planar distributed sources for radiating transient pulses. Under suitable approximations, this can be analyzed as what is referred to as an aperture antenna [4]. By assuming an appropriate distribution of the tangential electric field on the aperture the fields throughout space are expressible by integrals over the aperture.

In order to simplify the results and maximize the fields one can restrict consideration to what is called a focused aperture [4]. This separates the dependence of the fields at the observer into separate spatial and frequency terms. This also allows for a convenient representation in time domain.

After developing the general theory, explicit integrals are given for the spatial coefficients for the two frequency terms for the electric field and the three frequency terms for the magnetic field. Then the case of a circular aperture, focused on the axis of symmetry, with a uniform tangential electric field is considered, resulting in closed-form expressions valid at any distance from the aperture.

2. ELECTROMAGNETIC FIELDS FROM ARBITRARY APERTURE FIELDS IN A PLANE

The basic equations for electromagnetic fields from those on a half plane ($z = 0$) are well know [3, 4]. In particular let us consider that the tangential electric field on the source plane S' (i.e., $z' = 0$) is specified. Let the $\vec{r}\,'$ coordinates be those on S', i.e.,

$$\vec{r}\,' = (x', y', 0) \tag{1}$$

Furthermore, let $\vec{r}$ be the coordinates of the fields away from S' as

$$\vec{r} = (x, y, z) \tag{2}$$

Then as in Figure 1, we have some assumed tangential field distribution on S' as

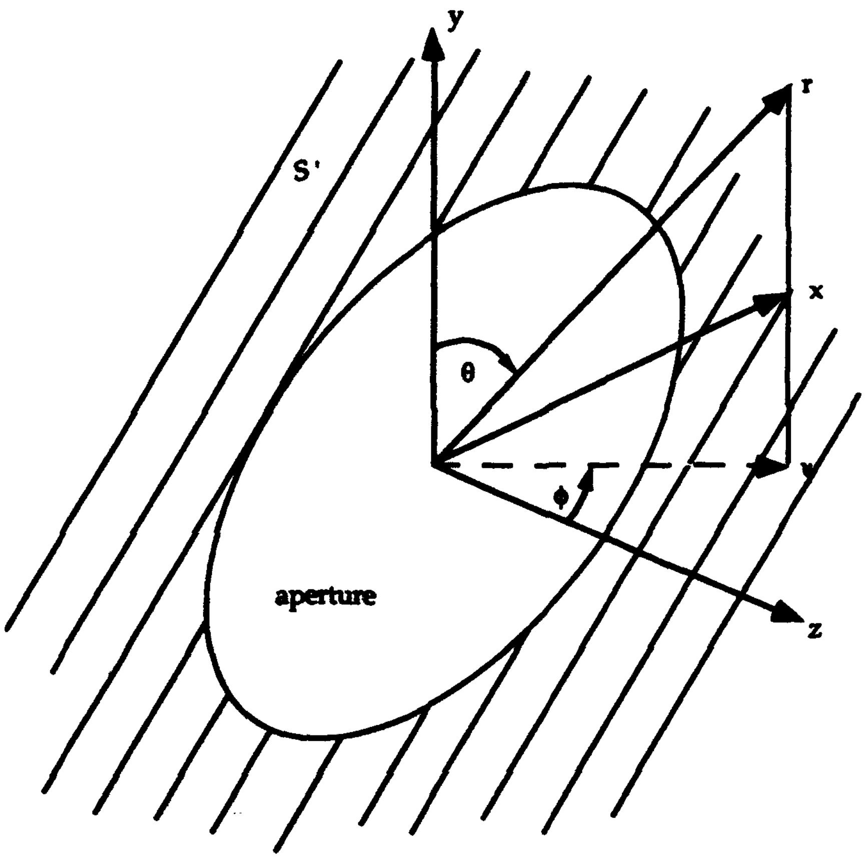
$$\vec{E}_t(x', y', 0; t) = \left(E_{x'}(x', y', 0; t), E_{y'}(x', y', 0; t), 0\right) \equiv \vec{E}_t(x', y'; t) \tag{3}$$

In addition, there is a z component of the electric field on S' which, however, does not appear as a source in the traditional equations.

Now at some position $\vec{r}$ we wish to compute the fields. Define

$$\begin{aligned}
R &\equiv \left|\vec{r} - \vec{r}'\right| = \left[(x - x')^2 + (y - y')^2 + z^2\right]^{1/2} \\
\vec{1}_R &\equiv \frac{x - x'}{R}\,\vec{1}_x + \frac{y - y'}{R}\,\vec{1}_y + \frac{z - z'}{R}\,\vec{1}_z \\
s &\equiv \text{Laplace-transform (two-sided) variable} \\
&\equiv \text{complex frequency} \\
\gamma &\equiv \frac{s}{c} \equiv \text{free-space propagation constant} \\
c &\equiv \frac{1}{\sqrt{\mu_0 \varepsilon_0}} \equiv \text{speed of light} \\
Z_0 &\equiv \sqrt{\frac{\mu_0}{\varepsilon_0}} \equiv \text{wave impedance of free space}
\end{aligned} \tag{4}$$

Our half space of interest ($z > 0$) is considered free space with zero conductivity, permittivity ε_0, and permeability μ_0.



S′ is defined by $\vec{r}' = (x', y', 0)$.

Figure 1. Electromagnetic fields from a source plane S'.

The electromagnetic fields are computed as from an equivalent set of magnetic current sources on S' which give a vector potential [1, 3].

$$\vec{\tilde{A}}(\vec{r},s) = \frac{-1}{2\pi s}\int_{S'} \frac{\gamma R+1}{R^2}\left[\left[\vec{1}_z \times \vec{\tilde{E}}'_t(x', y'; s)\right]\times \vec{1}_R\right] e^{-\gamma R} dS' \tag{5}$$

The scalar potential of a magnetic current distribution being zero, we have

$$\vec{E} = \frac{-\partial \vec{A}}{\partial t}, \ \vec{\tilde{E}} = -s\vec{\tilde{A}}, \ \vec{B} = \mu_0 \vec{H} = \nabla \times \vec{A} \tag{6}$$

Then we have

$$\tilde{\vec{E}}(\vec{r},s) = \frac{1}{2\pi}\int_{S'}\frac{\gamma R+1}{R^2}\left[\left[\vec{1}_z \times \tilde{\vec{E}}'_t(x',y';s)\right]\times \vec{1}_R\right] e^{-\gamma R}dS'$$

$$\tilde{\vec{H}}(\vec{r},s) = \frac{-1}{2\pi\mu_o s}\left\{\nabla \times \int_{S'}\frac{\gamma R+1}{R^2}\left[\left[\vec{1}_z \times \tilde{\vec{E}}'_t(x', y';s)\right]\times \vec{1}_R\right] e^{-\gamma R}dS'\right\} \tag{7}$$

In component form, we have for the electric field

$$\tilde{E}_{\substack{x\\y}}(\vec{r},s) = \frac{1}{2\pi}\int_{S'}\frac{\gamma R+1}{R^2}\cdot\frac{z}{R}\; e^{-\gamma R}\,\tilde{E}'_{\substack{x\\y}}(x', y';s)\; dS'$$

$$\tilde{E}_z(\vec{r},s) = \frac{1}{2\pi}\int_{S'}\frac{\gamma R+1}{R^2}e^{-\gamma R}\left[\frac{x - x'}{R}\;\vec{E}'_x(x', y';s) + \frac{y - y'}{R}\;\tilde{E}'_y(x', y';s)\right]dS' \tag{8}$$

The magnetic field components are somewhat more complicated as

$$\tilde{H}_x(\vec{r},s) = \frac{1}{2\pi\mu_o s}\int_{S'}\frac{e^{-\gamma R}}{R^3}\left\{\frac{-(x - x')(y - y')}{R^2}\left[(\gamma R)^2 + 3\gamma R + 3\right]\tilde{E}'_x(x', y';s)\right.$$

$$\left. + \left[2\gamma R + 2 - \frac{(y - y')^2 + z^2}{R^2}\left[(\gamma R)^2 + 3\gamma R + 3\right]\right]\tilde{E}'_y(x', y';s)\right\}dS'$$

$$\tilde{H}_y(\vec{r},s) = \frac{1}{2\pi\mu_o s}\int_{S'}\frac{e^{-\gamma R}}{R^3}\left\{\frac{(x - x')(y - y')}{R^2}\left[(\gamma R)^2 + 3\gamma R + 3\right]\tilde{E}'_y(x', y';s)\right.$$

$$\left. - \left[2\gamma R + 2 - \frac{(x - x')^2 + z^2}{R^2}\left[(\gamma R)^2 + 3\gamma R + 3\right]\right]\tilde{E}'_x(x', y';s)\right\}dS'$$

$$\tilde{H}_z(\vec{r},s) = \frac{1}{2\pi\mu_o s}\int_{S'}\frac{e^{-\gamma R}}{R^3}\left\{\frac{-(y - y')z}{R^2}\left[(\gamma R)^2 + 3\gamma R + 3\right]\tilde{E}'_x(x', y';s)\right. \tag{9}$$

$$\left. + \frac{(x - x')z}{R^2}\left[(\gamma R)^2 + 3\gamma R + 3\right]\tilde{E}'_y(x', y';s)\right\}dS'$$

Note that one can think of the fields from an aperture as those from an array of magnetic dipoles. The electric field has R^{-1} and R^{-2} contributions and the magnetic field has R^{-1}, R^{-2}, and R^{-3} contributions.

In time domain one also has explicit formulas by making the association.

$$
\begin{array}{rcl}
\text{Laplace (complex frequency) Domain} & \leftrightarrow & \text{Time Domain} \\
e^{-\gamma R}\ \vec{E}'_t(x', y'; s) & \leftrightarrow & \vec{E}'_t\left(x', y'; t - \frac{R}{c}\right) \\
s & \leftrightarrow & \frac{\partial}{\partial t} \\
s^{-1} & \leftrightarrow & \int dt
\end{array}
\qquad (10)
$$

Substituting the time-domain forms in (7) through (9) these equations are all converted to explicit time-domain formulae for the fields in terms of time-domain aperture fields.

3. FOCUSED APERTURE

As discussed in various texts [4], one can have a focused aperture. By this we mean that the phase of the tangential electric field (source) on S' is adjusted such that at some observer at $\vec{r} = \vec{r}_o$ the signals from each elementary position on S' all arrive with the same phase. Looking at (7) it is the factor $e^{-\gamma R}\ \vec{E}'_t$, which we need to control. While $\vec{E}'_t$ is only a function of the source coordinates, $e^{-\gamma R}$ is a function of both source and field coordinates.

Considering a fixed observer position at $\vec{r} = \vec{r}_o$ then we have

$$R = \left| \vec{r}_o - \vec{r} \right|, \quad R_o = \left| \vec{r}_o \right| = r_o \tag{11}$$

Let us now constrain at $\vec{r}_o$

$$e^{-\gamma R} \tilde{\vec{E}}'_t(x', y'; s) = e^{-\gamma R_o} E_o \tilde{f}(s) \, \vec{g}(x', y') \tag{12}$$

With this choice, then at $\vec{r}_o$ all sources (or equivalent elementary magnetic dipoles) have the same waveform $f(t)$ arriving at the same time at $\vec{r}_o$. The variation of the source amplitude over S' is constrained by a separate factor $\vec{g}(x', y')$ (real) with an overall scale factor E_0 (real).

Substituting (12) into (7) through (9) gives the fields at the focus $\vec{r}_o$. Note that in the magnetic field equation in (7) the curl operator is with respect to $\vec{r}$, not $\vec{r}_o$. In terms of the components the electric field at $\vec{r}_o$ is

$$\tilde{E}_{\substack{x\\y}}(\vec{r}_o, s) = \frac{e^{-\gamma R_o}}{2\pi} E_o \tilde{f}(s) \int_{S'} \frac{\gamma R + 1}{R^2} \frac{z_o}{R} \, g_{\substack{x\\y}}(x', y') \, dS'$$

$$\tilde{E}_z(\vec{r}_o, s) = \frac{e^{-\gamma R_o}}{2\pi} E_o \tilde{f}(s) \int_{S'} \frac{\gamma R + 1}{R^2} \left[\frac{x_o - x'}{R} g_x(x', y') + \frac{y_o - y'}{R} g_y(x', y') \right] dS' \tag{13}$$

Similarly, the magnetic field is

$$\tilde{H}_x(\vec{r}_o, s) = \frac{e^{-\gamma R_o}}{2\pi\mu_o s} E_o \tilde{f}(s) \int_{S'} \left\{ \frac{-(x_o - x')(y_o - y')}{R^5} \left[(\gamma R)^2 + 3\gamma R + 3 \right] g_x(x', y') \right.$$

$$\left. + \left[\frac{2\gamma R + 2}{R^3} - \frac{(y_o - y')^2 + z_o^2}{R^5} \left[(\gamma R)^2 + 3\gamma R + 3 \right] \right] g_y(x', y') \right\} dS'$$

$$\tilde{H}_y(\vec{r}_o,s) = \frac{e^{-\gamma R_o}}{2\pi\mu_o s} E_o \tilde{f}(s) \int_{S'} \left\{ \frac{(x_o - x')(y_o - y')}{R^5} \left[(\gamma R)^2 + 3\gamma R + 3 \right] g_y(x', y') \right.$$

$$\left. - \left[\frac{2\gamma R + 2}{R^3} - \frac{(x_o - x')^2 + z_o^2}{R^5} \left[(\gamma R)^2 + 3\gamma R + 3 \right] \right] g_x(x', y') \right\} dS'$$

$$\tilde{H}_z(\vec{r}_o,s) = \frac{e^{-\gamma R_o}}{2\pi\mu_o s} E_o \tilde{f}(s) \int_{S'} \left\{ \frac{-(y_o - y')z_o}{R^5} \left[(\gamma R)^2 + 3\gamma R + 3 \right] g_x(x', y') \right.$$

$$\left. + \frac{(x_o - x')\,z_o}{R^5} \left[(\gamma R)^2 + 3\gamma R + 3 \right] g_y(x', y') \right\} dS' \tag{14}$$

After allowing for the delay term $e^{-\gamma R_o}$ and the source (tangential E) waveform $\tilde{f}(s)$, note that the electric field has terms proportional to s^1 and s^0, while the magnetic field has terms proportional to s^1, s^0 and s^{-1}. Then let us write for the electric field

$$\vec{E}_x(\vec{r}_o,s) = E_o e^{-\gamma R_o} \tilde{f}(s) \left\{ \alpha_{x,x}^{(1)} \frac{a}{R_o} \frac{as}{c} + \alpha_{x,x}^{(2)} \frac{a^2}{R_o^2} \right\}$$

$$\vec{E}_y(\vec{r}_o,s) = E_o e^{-\gamma R_o} \tilde{f}(s) \left\{ \alpha_{y,y}^{(1)} \frac{a}{R_o} \frac{as}{c} + \alpha_{y,y}^{(2)} \frac{a^2}{R_o^2} \right\} \tag{15}$$

$$\vec{E}_z(\vec{r}_o,s) = E_o e^{-\gamma R_o} \tilde{f}(s) \left\{ \left[\alpha_{z,x}^{(1)} + \alpha_{z,y}^{(1)} \right] \frac{a}{R_o} \frac{as}{c} + \left[\alpha_{z,x}^{(2)} + \alpha_{z,y}^{(2)} \right] \frac{a^2}{R_o^2} \right\}$$

Similarly, for the magnetic field

$$Z_o\,\tilde{H}_x(\vec{r}_o,s) = E_o e^{-\gamma R_o}\,\tilde{f}(s)\left\{\beta^{(1)}_{x,x}\frac{a}{R_o}\frac{as}{c} + \beta^{(2)}_{x,x}\frac{a^2}{R_o^2} + \beta^{(3)}_{x,x}\frac{a^3}{R_o^3}\frac{c}{as}\right.$$

$$\left. + \beta^{(1)}_{x,y}\frac{a}{R_o}\frac{as}{c} + \beta^{(2)}_{x,y}\frac{a^2}{R_o^2} + \beta^{(3)}_{x,y}\frac{a^3}{R_o^3}\frac{c}{as}\right\}$$

$$Z_o\,\tilde{H}_y(\vec{r}_o,s) = E_o e^{-\gamma R_o}\,\tilde{f}(s)\left\{\beta^{(1)}_{y,x}\frac{a}{R_o}\frac{as}{c} + \beta^{(2)}_{y,x}\frac{a^2}{R_o^2} + \beta^{(3)}_{y,x}\frac{a^3}{R_o^3}\frac{c}{as}\right.$$

$$\left. + \beta^{(1)}_{y,y}\frac{a}{R_o}\frac{as}{c} + \beta^{(2)}_{y,y}\frac{a^2}{R_o^2} + \beta^{(3)}_{y,y}\frac{a^3}{R_o^3}\frac{c}{as}\right\}$$

$$Z_o\,\tilde{H}_z(\vec{r}_o,s) = E_o e^{-\gamma R_o}\,\tilde{f}(s)\left\{\beta^{(1)}_{z,x}\frac{a}{R_o}\frac{as}{c} + \beta^{(2)}_{z,x}\frac{a^2}{R_o^2} + \beta^{(3)}_{z,x}\frac{a^3}{R_o^3}\frac{c}{as}\right. \tag{16}$$

$$\left. + \beta^{(1)}_{z,y}\frac{a}{R_o}\frac{as}{c} + \beta^{(2)}_{z,y}\frac{a^2}{R_o^2} + \beta^{(3)}_{z,y}\frac{a^3}{R_o^3}\frac{c}{as}\right\}$$

where

$$a \equiv \text{some characteristic dimension of the aperture} \tag{17}$$

In this form the α and β coefficients are frequency independent and dimensionless.

The above results are also directly expressable in time domain for the electric field as

$$E_x(\vec{r}_o,t) = E_o\left\{\alpha^{(2)}_{x,x}\frac{a^2}{R_o c}\frac{\partial}{\partial t}f\left(t-\frac{R_o}{c}\right) + \alpha^{(2)}_{x,x}\frac{a^2}{R_o^2}\,f\left(t-\frac{R_o}{c}\right)\right\}$$

$$E_y(\vec{r}_o,t) = E_o\left\{\alpha^{(1)}_{y,y}\frac{a^2}{R_o c}\frac{\partial}{\partial t}f\left(t-\frac{R_o}{c}\right) + \alpha^{(2)}_{y,y}\frac{a^2}{R_o^2}\,f\left(t-\frac{R_o}{c}\right)\right\} \tag{18}$$

$$E_z(\vec{r}_o,t) = E_o\left\{\left[\alpha^{(1)}_{z,x} + \alpha^{(1)}_{z,y}\right]\frac{a^2}{R_o c}\frac{\partial}{\partial t}f\left(t-\frac{R_o}{c}\right) + \left[\alpha^{(2)}_{z,x} + \alpha^{(1)}_{z,x}\right]\frac{a^2}{R_o^2}f\left(t-\frac{R_o}{c}\right)\right\}$$

In time domain the magnetic field is

$$Z_o H_x(\vec{r}_o,t) = E_o \left\{ \left[\beta^{(1)}_{x,x} + \beta^{(1)}_{x,y} \right] \frac{a^2}{R_o c} \frac{\partial}{\partial t} f\left(t - \frac{R_o}{c}\right) \right.$$

$$\left. + \left[\beta^{(2)}_{x,x} + \beta^{(2)}_{x,y} \right] \frac{a^2}{R_o^2} f\left(t - \frac{R_o}{c}\right) + \left[\beta^{(3)}_{x,x} + \beta^{(3)}_{x,y} \right] \frac{a^2 c}{R_o^3} \int_{-\infty}^{t} f\left(t' - \frac{R_o}{c}\right) dt' \right\}$$

$$Z_o H_y(\vec{r}_o,t) = E_o \left\{ \left[\beta^{(1)}_{y,x} + \beta^{(1)}_{y,y} \right] \frac{a^2}{R_o c} \frac{\partial}{\partial t} f\left(t - \frac{R_o}{c}\right) \right.$$

$$\left. + \left[\beta^{(2)}_{y,x} + \beta^{(2)}_{y,y} \right] \frac{a^2}{R_o^2} f\left(t - \frac{R_o}{c}\right) + \left[\beta^{(3)}_{y,x} + \beta^{(3)}_{y,y} \right] \frac{a^2 c}{R_o^3} \int_{-\infty}^{t} f\left(t' - \frac{R_o}{c}\right) dt' \right\}$$

$$Z_o H_z(\vec{r}_o,t) = E_o \left\{ \left[\beta^{(1)}_{z,x} + \beta^{(1)}_{z,y} \right] \frac{a^2}{R_o c} \frac{\partial}{\partial t} f\left(t - \frac{R_o}{c}\right) \right.$$

$$\left. + \left[\beta^{(2)}_{z,x} + \beta^{(2)}_{z,y} \right] \frac{a^2}{R_o^2} f\left(t - \frac{R_o}{c}\right) + \left[\beta^{(3)}_{z,x} + \beta^{(3)}_{z,y} \right] \frac{a^2 c}{R_o^3} \int_{-\infty}^{t} f\left(t' - \frac{R_o}{c}\right) dt' \right\} \tag{19}$$

Note for the time integrals the excitation waveform is assumed to be identically zero before some turn-on time. In (18) and (19) the utility of focusing at $\vec{r}_o$ is indicated by the simple form the results take in time domain. If the aperture sources did not all reach $\vec{r}_o$ at the same time the three time-domain terms (derivative waveform, waveform, and integral waveform) would be smeared out. This is another way of considering that the fields are maximized at $\vec{r}_o$ by focusing.

Now we have the coefficients which depend on $\vec{r}_o$ and the spatial distribution on the source. The coefficients of the derivative terms for the electric field are

$$\alpha^{(1)}_{x,x} = \frac{1}{2\pi}\frac{R_0 z_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-1} g_x(x', y')\, dS'$$

$$\alpha^{(1)}_{y,y} = \frac{1}{2\pi}\frac{R_0 z_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-1} g_y(x', y')\, dS'$$

$$\alpha^{(1)}_{z,x} = \frac{1}{2\pi}\frac{R_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-1} (x_o - x')\, g_x(x', y')\, dS' \tag{20}$$

$$\alpha^{(1)}_{z,y} = \frac{1}{2\pi}\frac{R_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-1} (y_o - y')\, g_y(x', y')\, dS'$$

The coefficients of the terms proportional to the source waveform for the electric field are

$$\alpha^{(2)}_{x,x} = \frac{1}{2\pi}\frac{R_0^2 z_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2} g_x(x', y')\, dS'$$

$$\alpha^{(2)}_{y,y} = \frac{1}{2\pi}\frac{R_0^2 z_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2} g_y(x', y')\, dS'$$

$$\alpha^{(2)}_{z,x} = \frac{1}{2\pi}\frac{R_0^2}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2} (x_o - x')\, g_x(x', y')\, dS' \tag{21}$$

$$\alpha^{(2)}_{z,y} = \frac{1}{2\pi}\frac{R_0^2}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2} (y_o - y')\, g_y(x', y')\, dS'$$

The coefficients of the derivative terms for the magnetic field are

$$\beta^{(1)}_{x,x} = \frac{-1}{2\pi}\frac{R_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2} (x_o - x')(y_o - y')\, g_x(x', y')\, dS'$$

$$\beta^{(1)}_{x,y} = \frac{-1}{2\pi}\frac{R_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2} \left[(y_o - y')^2 + z_o^2\right] g_y(x', y')\, dS'$$

$$\beta^{(1)}_{y,x} = \frac{1}{2\pi}\frac{R_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2} \left[(x_o - x')^2 + z_o^2\right] g_x(x', y')\, dS'$$

$$\beta^{(1)}_{y,y} = \frac{1}{2\pi}\frac{R_o}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2}(x_o - x')(y_o - y')\; g_y(x', y')\,dS'$$

$$\beta^{(1)}_{z,x} = \frac{-1}{2\pi}\frac{R_o}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2}(y_o - y')\,z_o\, g_x(x', y')\,dS'$$

$$\beta^{(1)}_{z,y} = \frac{1}{2\pi}\frac{R_o}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2}(x_o - x')\,z_o\, g_y(x', y')\,dS' \tag{22}$$

The coefficients of the terms proportional to the source waveform for the magnetic field are

$$\beta^{(2)}_{x,x} = \frac{-3}{2\pi}\frac{R_o^2}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-2}(x_o - x')(y_o - y')\; g_x(x', y')\,dS'$$

$$\beta^{(2)}_{x,y} = \frac{1}{2\pi}\frac{R_o^2}{a^2}\int_{S'}\Big\{2\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-1}$$

$$- 3\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-2}\left[(y_o - y')^2 + z_o^2\right]\Big\}\; g_y(x', y')\,dS'$$

$$\beta^{(2)}_{y,x} = \frac{1}{2\pi}\frac{R_o^2}{a^2}\int_{S'}\Big\{-2\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-1}$$

$$+ 3\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-2}\left[(x_o - x')^2 + z_o^2\right]\Big\}\; g_x(x', y')\,dS'$$

$$\beta^{(2)}_{y,y} = \frac{3}{2\pi}\frac{R_o^2}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-2}(x_o - x')(y_o - y')\; g_y(x', y')\,dS'$$

$$\beta^{(2)}_{z,x} = \frac{-3}{2\pi}\frac{R_o^2 z_o}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-2}(y_o - y')\; g_x(x', y')\,dS'$$

$$\beta^{(2)}_{z,y} = \frac{3}{2\pi}\frac{R_o^2 z_o}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-2}(x_o - x')\; g_y(x', y')\,dS' \tag{23}$$

The coefficients of the integral terms for the magnetic field are

$$\beta^{(3)}_{x,x} = \frac{-3}{2\pi}\frac{R_0^3}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-5/2}(x_o - x')(y_o - y')\; g_x(x', y')\,dS'$$

$$\beta^{(3)}_{x,y} = \frac{1}{2\pi}\frac{R_0^3}{a^2}\int_{S'}\left\{2\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2}\right.$$

$$\left. - 3\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-5/2}\left[(y_o - y')^2 + z_o^2\right]\right\} g_y(x', y')\,dS'$$

$$\beta^{(3)}_{y,x} = \frac{1}{2\pi}\frac{R_0^2}{a^2}\int_{S'}\left\{-2\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-3/2}\right.$$

$$\left. + 3\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-5/2}\left[(x_o - x')^2 + z_o^2\right]\right\} g_x(x', y')\,dS'$$

$$\beta^{(3)}_{y,y} = \frac{3}{2\pi}\frac{R_0^3}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-5/2}(x_o - x')(y_o - y')\; g_y(x', y')\,dS'$$

$$\beta^{(3)}_{z,x} = \frac{-3}{2\pi}\frac{R_0^3 z_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-5/2}(y_o - y')\; g_x(x', y')\,dS'$$

$$\beta^{(3)}_{z,y} = \frac{3}{2\pi}\frac{R_0^3 z_0}{a^2}\int_{S'}\left[(x_o - x')^2 + (y_o - y')^2 + z_o^2\right]^{-5/2}(x_o - x')\; g_y(x', y')\,dS' \tag{24}$$

4. CASE OF FOCUS AT LARGE DISTANCE FROM APERTURE

Now let $R_0 / a \rightarrow \infty$ where a is some characteristic dimension of the aperture. Specifically let all source fields be zero outside of some radius on S' given by some constant times a. For this purpose we have coordinates (cartesian, cylindrical, spherical) as

$$
\begin{aligned}
x &= \Psi \cos(\phi), & y &= \Psi \sin(\phi) \\
z &= r \cos(\theta), & \Psi &= r \sin(\theta) \\
x &= r \sin(\theta)\ \cos(\phi), & y &= r \sin(\theta)\ \sin(\phi)
\end{aligned}
\qquad (25)
$$

These are subscripted with 0 for the observer at $\vec{r} = \vec{r}_o$.

Considering that θ_0 and ϕ_0 are fixed as $R_o / a \to \infty$, we have to order $(a/R_0)^{-1}$ for the electric coefficients in (20) and (21).

$$\alpha^{(1)}_{x,x} = \alpha^{(2)}_{x,x} = \frac{1}{2\pi} \cos(\theta_o) \frac{1}{a^2} \int_{S'} g_x(x', y')\ dS'$$

$$\alpha^{(1)}_{y,y} = \alpha^{(2)}_{x,x} = \frac{1}{2\pi} \cos(\theta_o) \frac{1}{a^2} \int_{S'} g_y(x', y')\ dS'$$

$$\alpha^{(1)}_{z,x} = \alpha^{(2)}_{z,x} = \frac{1}{2\pi} \sin(\theta_o)\ \cos(\phi_o) \frac{1}{a^2} \int_{S'} g_x(x', y')\ dS'$$

$$\alpha^{(1)}_{z,y} = \alpha^{(2)}_{z,x} = \frac{1}{2\pi} \sin(\theta_o)\ \sin(\phi_o) \frac{1}{a^2} \int_{S'} g_y(x', y')\ dS' \qquad (26)$$

as $\frac{a}{R_o} \to 0$ to $O\left(\frac{a}{R_o}\right)$

Similarly for the magnetic coefficients in (22) through (24)

$$\beta^{(1)}_{x,x} = \frac{1}{3}\ \beta^{(2)}_{x,x} = \frac{1}{3}\ \beta^{(3)}_{x,x} = \frac{-1}{2\pi} \sin^2(\theta_o)\ \cos(\phi_o)\ \sin(\phi_o)\ \frac{1}{a^2} \int_{S'} g_x(x', y')\ dS'$$

$$\beta^{(1)}_{x,y} = \frac{-1}{2\pi} \left[\cos^2(\theta_o) + \sin^2(\theta_o)\ \sin^2(\phi_o)\right] \frac{1}{a^2} \int_{S'} g_y(x', y')\ dS'$$

$$\beta^{(2)}_{x,y} = \frac{1}{2\pi} \left[2 - 3\left[\cos^2(\theta_o) + \sin^2(\theta_o)\ \sin^2(\phi_o)\right]\right] \frac{1}{a^2} \int_{S'} g_y(x', y')\ dS'$$

$$\beta^{(3)}_{x,y} = \frac{1}{2\pi}\left[2-3\left[\cos^2(\theta_o) + \sin^2(\theta_o)\ \sin^2(\phi_o)\right]\right]\frac{1}{a^2}\int_{S'} g_y(x', y')\ dS'$$

$$\beta^{(1)}_{y,x} = \frac{1}{2\pi}\left[\cos^2(\theta_o) + \sin^2(\theta_o)\ \cos^2(\phi_o)\right]\frac{1}{a^2}\int_{S'} g_x(x', y')\ dS'$$

$$\beta^{(2)}_{y,x} = \frac{1}{2\pi}\left[-2+3\left[\cos^2(\theta_o) + \sin^2(\theta_o)\ \cos^2(\phi_o)\right]\right]\frac{1}{a^2}\int_{S'} g_x(x', y')\ dS'$$

$$\beta^{(3)}_{y,x} = \frac{-1}{2\pi}\left[-2+3\left[\cos^2(\theta_o) + \sin^2(\theta_o)\ \cos^2(\phi_o)\right]\right]\frac{1}{a^2}\int_{S'} g_x(x', y')\ dS'$$

$$\beta^{(1)}_{z,x} = \frac{1}{3}\ \beta^{(2)}_{z,x} = \frac{1}{3}\ \beta^{(3)}_{z,x} = \frac{-1}{2\pi}\cos(\theta_o)\ \sin(\theta_o)\ \sin(\phi_o)\ \frac{1}{a^2}\int_{S'} g_x(x', y')\ dS'$$

$$\beta^{(1)}_{z,y} = \frac{1}{3}\ \beta^{(2)}_{z,y} = \frac{1}{3}\ \beta^{(3)}_{z,y} = \frac{1}{2\pi}\cos(\theta_o)\ \sin(\theta_o)\ \cos(\phi_o)\ \frac{1}{a^2}\int_{S'} g_y(x', y')\ dS' \quad (27)$$

as $\frac{a}{R_o} \to 0$ to $O\left(\frac{a}{R_o}\right)$

Note the common factors in the above. If we choose as a special case a uniform illumination (after focusing) of a circular aperture of radius a, with a single polarization in the x direction, we have

$$g_x(x', y') = \begin{cases} 1 \text{ for } 0 \le \Psi' < a \\ 0 \text{ for } \Psi' > a \end{cases}$$

$$g_y(x', y') = 0$$

$$\frac{1}{a^2}\int_{S'} g_x(x', y')\ dS' = \pi \qquad (28)$$

$$\frac{1}{a^2}\int_{S'} g_y(x', y')\ dS' = 0$$

These can be substituted in (26) and (27) to give the usual results for a uniformly illuminated circular aperture focused at ∞.

In a more general sense, let the aperture be of area A with uniform illumination g_x. Then

$$\int_{S'} g_x(x', y')\, dS' = A \tag{29}$$

which is a well-known result for uniformly illuminated apertures focused at infinity.

5. CIRCULAR APERTURE WITH UNIFORM TANGENTIAL ELECTRIC FIELD FOCUSED ON AXIS

Now consider the special case of a circular aperture with uniform illumination (tangential electric field)

$$\begin{aligned} g_x(x', y') &\equiv \begin{cases} 1 & \text{for } 0 \leq \psi' \leq a \\ 0 & \text{for } \psi' > a \end{cases} \\ g_y(x', y') &\equiv 0 \end{aligned} \tag{30}$$

Furthermore, specialize the case to an observer at a distance z_0 on the z axis.

The non-zero electric-field coefficients are

$$\begin{aligned} \alpha_{x,x}^{(1)} &= \frac{1}{2\pi} \frac{z_0^2}{a^2} \int_{S'} \left[x'^2 + y'^2 + z_0^2 \right]^{-1} g_x(x', y')\, dS' \\ &= \frac{1}{2\pi} \frac{z_0^2}{a^2} \int_0^a \int_0^{2\pi} \left[\Psi'^2 + z_0^2 \right]^{-1} \Psi'\, d\phi'\, d\Psi' \\ &= \frac{z_0^2}{a^2} \int_0^a \left[\Psi'^2 + z_0^2 \right]^{-1} \Psi'\, d\Psi' \end{aligned}$$

$$= \frac{z_0^2}{2a^2}\int_0^{a^2}\left[v + z_0^2\right]^{-1} dv$$

$$= \frac{z_0^2}{2a^2}\,\ell n\left[1 + \frac{a^2}{z_0^2}\right]$$

$$\alpha^{(2)}_{x,x} = \frac{1}{2\pi}\,\frac{z_0^3}{a^2}\int_{S'}\left[x'^2 + y'^2 + z_0^2\right]^{-3/2} g_x(x', y')\, dS'$$

$$= \frac{1}{2\pi}\,\frac{z_0^3}{a^2}\int_0^{a}\int_0^{2\pi}\left[\Psi'^2 + z_0^2\right]^{-3/2}\Psi'\, d\phi'\, d\Psi'$$

$$= \frac{z_0^3}{a^2}\int_0^{a}\left[\Psi'^2 + z_0^2\right]^{-3/2}\Psi'\, d\Psi'$$

$$= \frac{z_0^3}{2a^2}\int_0^{a^2}\left[v + z_0^2\right]^{-3/2} dv$$

$$= \frac{-z_0^3}{a^2}\left\{\left[a^2 + z_0^2\right]^{-1/2} - z_0^{-1}\right\} \qquad (31)$$

$$= \frac{z_0^2}{a^2}\left\{1 - \left[1 + \frac{a^2}{z_0^2}\right]^{-1/2}\right\}$$

Substituting these results in (18) we have

$$E_x(z_0\,\vec{1}_z, t) = E_o\left\{\frac{z_o}{2c}\,\ell n\left[1 + \frac{a^2}{z_o^2}\right]\frac{\partial}{\partial t} f\left(t - \frac{z_o}{c}\right) + \left\{1 - \left[1 + \frac{a^2}{z_o^2}\right]^{-1/2}\right\} f\left(t - \frac{z_o}{c}\right)\right\} \qquad (32)$$

which is a quite compact result valid for $z_0 > 0$. This can be expanded for small a^2/z_o^2 to give

$$E_x(z_o \vec{1}_z, t) = E_o\left\{\frac{a^2}{2cz_o}\left[1 + O\left(\frac{a^2}{z_o^2}\right)\right]\frac{\partial}{\partial t} f\left(t - \frac{z_o}{c}\right)\right.$$

$$\left. + \frac{1}{2}\frac{a^2}{z_o^2}\left[1 + O\left(\frac{a^2}{z_o^2}\right)\right] f\left(t - \frac{z_o}{c}\right)\right\} \tag{33}$$

$$\text{as } \frac{a^2}{z_o^2} \to 0$$

This exhibits the usual far-field results (derivative term) proportional to (aperture area)/z_0. In addition for the term proportional to the excitation waveform there is a leading term proportional to (aperture area)/ z_o^2.

The non-zero magnetic-field coefficients are

$$\beta_{y,x}^{(1)} = \frac{1}{2\pi}\frac{z_o}{a^2}\int_{S'}\left[x'^2 + y'^2 + z_o^2\right]^{-3/2}\left[x'^2 + z_o^2\right] g_x(x', y')\, dS'$$

$$= \frac{1}{2\pi}\frac{z_o}{a^2}\int_o^a\int_o^{2\pi}\left[\Psi'^2 + z_o^2\right]^{-3/2}\left[\Psi'^2 \cos^2(\phi') + z_o^2\right]\Psi'\, d\phi'\, d\Psi'$$

$$= \frac{1}{2}\frac{z_o}{a^2}\int_o^a\left[\Psi'^2 + z_o^2\right]^{-3/2}\left[\Psi'^2 + 2z_o^2\right]\Psi'\, d\Psi'$$

$$= \frac{1}{4}\frac{z_o}{a^2}\int_o^{a^2}\left[v + z_o^2\right]^{-3/2}\left[v + 2z_o^2\right] dv$$

$$= \frac{1}{4}\frac{z_o}{a^2}\Big\{2\ \Big\{\big[a^2 + z_o^2\big]^{1/2} + z_o^2\ \big[a^2 + z_o^2\big]^{-1/2} - 2z_o\Big\}$$

$$- 4z_o^2\ \Big\{\big[a^2 + z_o^2\big]^{-1/2} - z_o^{-1}\Big\}\Big\}$$

$$= \frac{1}{2}\frac{z_o^2}{a^2}\left\{\left[1 + \frac{a^2}{z_o^2}\right]^{1/2} - \left[1 + \frac{a^2}{z_o^2}\right]^{-1/2}\right\}$$

$$\beta_{y,x}^{(2)} = \frac{1}{2\pi}\ \frac{z_o^2}{a^2}\int_{S'}\Big\{-2\big[x'^2 + y'^2 + z_o^2\big]^{-1}$$

$$+ 3\big[x'^2 + y'^2 + z_o^2\big]^{-2}\big[x'^2 + z_o^2\big]\Big\}g_x(x', y')\,dS'$$

$$= \frac{1}{2\pi}\ \frac{z_o^2}{a^2}\int_o^a\int_o^{2\pi}\left\{-2\left[\Psi'^2 + z_o^2\right]^{-1} + 3\left[\Psi'^2 + z_o^2\right]^{-2}\left[\Psi'^2\cos^2(\phi') + z_o^2\right]\right\}\Psi'\,d\phi'\,d\Psi'$$

$$= \frac{1}{2}\ \frac{z_o^2}{a^2}\int_o^a\left\{-4\left[\Psi'^2 + z_o^2\right]^{-1} + 3\left[\Psi'^2 + z_o^2\right]^{-2}\left[\Psi'^2 + 2z_o^2\right]\right\}\Psi'\,d\Psi'$$

$$= \frac{1}{4}\ \frac{z_o^2}{a^2}\int_o^{a^2}\left\{-4\left[\upsilon + z_o^2\right]^{-1} + 3\left[\upsilon + z_o^2\right]^{-2}\left[\upsilon + 2z_o^2\right]\right\}d\upsilon$$

$$= \frac{1}{4}\ \frac{z_o^2}{a^2}\Big\{-4\,\ell n\left[1 + \frac{a^2}{z_o^2}\right] + 3\ \Big\{\ell n\left[1 + \frac{a^2}{z_o^2}\right] + z_o^2\Big[\big[a^2 + z_o^2\big]^{-1} - z_o^{-2}\Big]\Big\}$$

$$- 6z_o^2\Big[\big[a^2 + z_o^2\big]^{-1} - z_o^{-2}\Big]\Big\}$$

$$= \frac{z_o^2}{a^2}\left\{\frac{-1}{4}\,\ell n\left[1 + \frac{a^2}{z_o^2}\right] - \frac{3}{4}\left[\left[1 + \frac{a^2}{z_o^2}\right]^{-1} - 1\right]\right\}$$

$$\beta_{y,x}^{(3)} = \frac{1}{2\pi}\,\frac{z_o^3}{a^2}\int_{S'}\left\{-2\left[x'^2 + y'^2 + z_o^2\right]^{-3/2}\right.$$

$$\left. + 3\left[x'^2 + y'^2 + z_o^2\right]^{-5/2}\left[x'^2 + z_o^2\right]\right\} g_x(x', y')\,dS'$$

$$= \frac{1}{2\pi}\,\frac{z_o^3}{a^2}\int_o^a\int_o^{2\pi}\left\{-2\left[\Psi'^2 + z_o^2\right]^{-3/2}\right.$$

$$\left. +3\left[\Psi'^2 + z_o^2\right]^{-5/2}\left[\Psi'^2\cos^2(\phi') + z_o^2\right]\right\}\Psi'\,d\phi'\,d\Psi'$$

$$= \frac{1}{2}\,\frac{z_o^3}{a^2}\int_o^a\left\{-4\left[\Psi'^2 + z_o^2\right]^{-3/2} + 3\left[\Psi'^2 + z_o^2\right]^{-5/2}\left[\Psi'^2 + 2z_o^2\right]\right\}\Psi'\,d\Psi'$$

$$= \frac{1}{4}\,\frac{z_o^3}{a^2}\int_o^{a^2}\left\{-4\left[\upsilon + z_o^2\right]^{-3/2} + 3\left[\upsilon + z_o^2\right]^{-5/2}\left[\upsilon + 2z_o^2\right]\right\}d\upsilon$$

$$= \frac{1}{4}\,\frac{z_o^3}{a^2}\left\{8\left\{\left[a^2 + z_o^2\right]^{-1/2} - z_o^{-1}\right\} + 6\left\{-\left[a^2 + z_o^2\right]^{-1/2} + z_o^{-1} + \frac{z_o^2}{3}\left[a^2 + z_o^2\right]^{-3/2}\right.\right.$$

$$\left.\left. - \frac{1}{3}z_o^{-1}\right\} - 4z_o^2\left\{\left[a^2 + z_o^2\right]^{-3/2} - z_o^{-3}\right\}\right\}$$

$$= \frac{z_o^2}{a^2}\left\{\frac{1}{2}\left[1 + \frac{a^2}{z_o^2}\right]^{-1/2} - \frac{1}{2}\left[1 + \frac{a^2}{z_o^2}\right]^{-3/2}\right\} \tag{34}$$

Substituting these results in (19) we have

$$Z_o\, H_y(z_o\, \vec{1}_z, t) = E_o \left\{ \frac{1}{2}\frac{z_o}{c} \left\{ \left[1+\frac{a^2}{z_o^2}\right]^{1/2} - \left[1+\frac{a^2}{z_o^2}\right]^{-1/2} \right\} \frac{\partial}{\partial t} f\left(t-\frac{z_o}{c}\right) \right.$$

$$+ \left\{ \frac{-1}{4}\, \ell n\left[1+\frac{a^2}{z_o^2}\right] - \frac{3}{4}\left[\left[1+\frac{a^2}{z_o^2}\right]^{-1} - 1\right]\right\} f\left(t-\frac{z_o}{c}\right) \tag{35}$$

$$\left. + \frac{c}{z_o}\left\{ \frac{1}{2}\left[1+\frac{a^2}{z_o^2}\right]^{-1/2} - \frac{1}{2}\left[1+\frac{a^2}{z_o^2}\right]^{-3/2} \right\} \int_{-\infty}^{t} f\left(t'-\frac{z_o}{c}\right) dt' \right\}$$

which again is valid for $z_0 > 0$. This can be expanded for small a^2/z_o^2 to give

$$Z_o\, H_y(z_o\, \vec{1}_z, t) = E_o \left\{ \frac{a^2}{2cz_o}\left[1+O\left(\frac{a^2}{z_o^2}\right)\right] \frac{\partial}{\partial t} f\left(t-\frac{z_o}{c}\right) \right.$$

$$+ \frac{1}{2}\frac{a^2}{z_o^2}\left[1+O\left(\frac{a^2}{z_o^2}\right)\right] f\left(t-\frac{z_o}{c}\right)$$

$$\left. + \frac{1}{2}\frac{ca^2}{z_o^3}\left[1+O\left(\frac{a^2}{z_o^2}\right)\right] \int_{-\infty}^{t} f\left(t'-\frac{z_o}{c}\right) dt' \right\} \tag{36}$$

as $\frac{a^2}{z_o^2} \to 0$

Note in the far field the leading terms (going as z_o^{-1}) agree for electric and magnetic fields. As one approaches the aperture so that a/z_0 becomes not small compared to 1, even the derivative terms for electric and magnetic fields have different coefficients (i.e., not related as in a plane wave).

6. SUMMARY

The case of Section 5 is not the only specific aperture distribution one could consider. Considering a focused aperture other distributions of aperture fields can be investigated and the coefficients in Section 3 determined. Perhaps some other interesting distributions will give closed-form results as in Section 5.

REFERENCES

1. Baum, C. E., (Dec 1968) "A Simplified Two-Dimensional Model for the Fields Above the Distributed-Source Surface Transmission Line", Sensor and Simulation Note 66.
2. Baum, C. E., (March 1970) "Some Characteristics of Planar Distributed Sources for Radiating Transient Pulses", Sensor and Simulation Note 100.
3. Smythe, W. R., (1968) Static and Dynamic Electricity, 3rd Edition, McGraw Hill.
4. Collin, R. E. and Zucker F. J. (eds.) (1969) Antenna Theory, McGraw Hill.

THE RADIATION PATTERN OF REFLECTOR IMPULSE RADIATING ANTENNAS: EARLY-TIME RESPONSE

Everett G. Farr
Farr Research

Carl E. Baum
Phillips Laboratory

Abstract

We generate here an approximation to the step response of reflector Impulse Radiating Antennas (IRAs) in the E- and H-planes. These step responses are then convolved with the derivative of the driving voltage to find the radiated fields. This allows a determination of the radiation pattern of reflector IRAs.

1. Introduction

Reflector Impulse Radiating Antennas (IRAs) consist of a paraboloidal reflector fed by a TEM feed (Figure 1). Such an arrangement creates a planar field distribution in an aperture plane. In order to calculate the field radiated by an IRA, one must understand the transient fields radiated from apertures. The transient radiation from a general planar aperture field was developed in [1]. This was further specialized to two-wire apertures in [2], but only on boresight. The off-boresight fields created by a two-wire aperture were never developed. We do so in this paper.

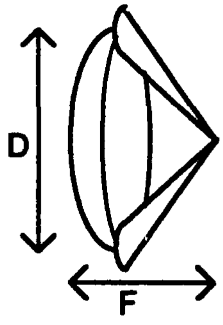


Figure 1. An Impulse Radiating Antenna.

The results we provide here are valid for a wide variety of practical cases. Although it is true that these results are valid only at high impedances, this is not very confining in practice, because at low impedances one incurs significant feed blockage anyway. Furthermore, our expressions are derived explicitly only for round wire feeds. But since the feed arms are relatively thin at high impedances, the specific shape of the feed should not matter very much. Thus, the results developed here are useful for other feed arm shapes, such as coplanar plates and curved plates [3].

We begin by providing a review of the static fields between two wires. This is the field we find in the aperture of a reflector IRA. Next, we radiate this field with a step-function driving voltage. Expressions are provided for the E-plane and H-plane fields (pattern) created when the aperture fields are turned on suddenly. Furthermore, we demonstrate that on boresight our results are consistent with earlier results calculated by contour integration. Finally, we convolve the step response with the derivative of a typical driving voltage, in order to get the response to a realistic driving voltage. With this information, we can calculate a gain and antenna pattern, according to the definitions of [4].

2. Static Fields in the Two-Wire Aperture

Let us review first the static fields in a two-wire aperture. We assume a Reflector IRA with a relatively high feed impedance, so we can ignore aperture blockage for early times. We can estimate the error incurred by ignoring aperture blockage from [3].

Recall that the effect of the paraboloidal reflector is to convert a spherical TEM wave into a (locally) planar TEM wave [5, Appendix A]. Thus, we must radiate an aperture field that turns on suddenly, as shown in Figure 2. We assume the tangential aperture field has the form

$$E_t(x,y,t) \;=\; E_{t_x} - jE_{t_y} \;=\; -\frac{V}{\Delta u}\frac{dw(\zeta)}{d\zeta}\,u(t) \tag{2.1}$$

where V is the voltage between the two wires, $u(t)$ is the Heaviside step function, and $w(\zeta)$ is the complex potential between the two wires described by

$$\begin{aligned} w(\zeta) &= u(\zeta) + j\,v(\zeta) \\ \zeta &= x + j\,y \\ u(\zeta)\ \&\ v(\zeta) &= \text{real functions} \end{aligned} \tag{2.2}$$

Furthermore, we restrict ourselves to round wires, which makes the math considerably simpler. For this case, the potential function is [6]

$$w(\zeta) \;=\; 2j\,\operatorname{arccot}(\zeta/a) \;=\; \ln\left(\frac{\zeta - ja}{\zeta + ja}\right) \tag{2.3}$$

where a is the aperture radius. This potential function is plotted in Figure 3. The real and imaginary parts of the potential can be separated as [10]

$$\begin{aligned} u(\zeta) &= \frac{1}{2}\ln\left[\frac{x^2+(a+y)^2}{x^2+(a-y)^2}\right] \\ v(\zeta) &= \arctan\left[\frac{2x}{x^2+y^2-a^2}\right] \end{aligned} \tag{2.4}$$

Finally, the impedance is

$$f_g \;=\; \frac{\Delta u}{\Delta v} \tag{2.5}$$

where Δu is the change in u from one conductor to the other, and Δv is the change in v as one encircles a conductor (in this case 2π).

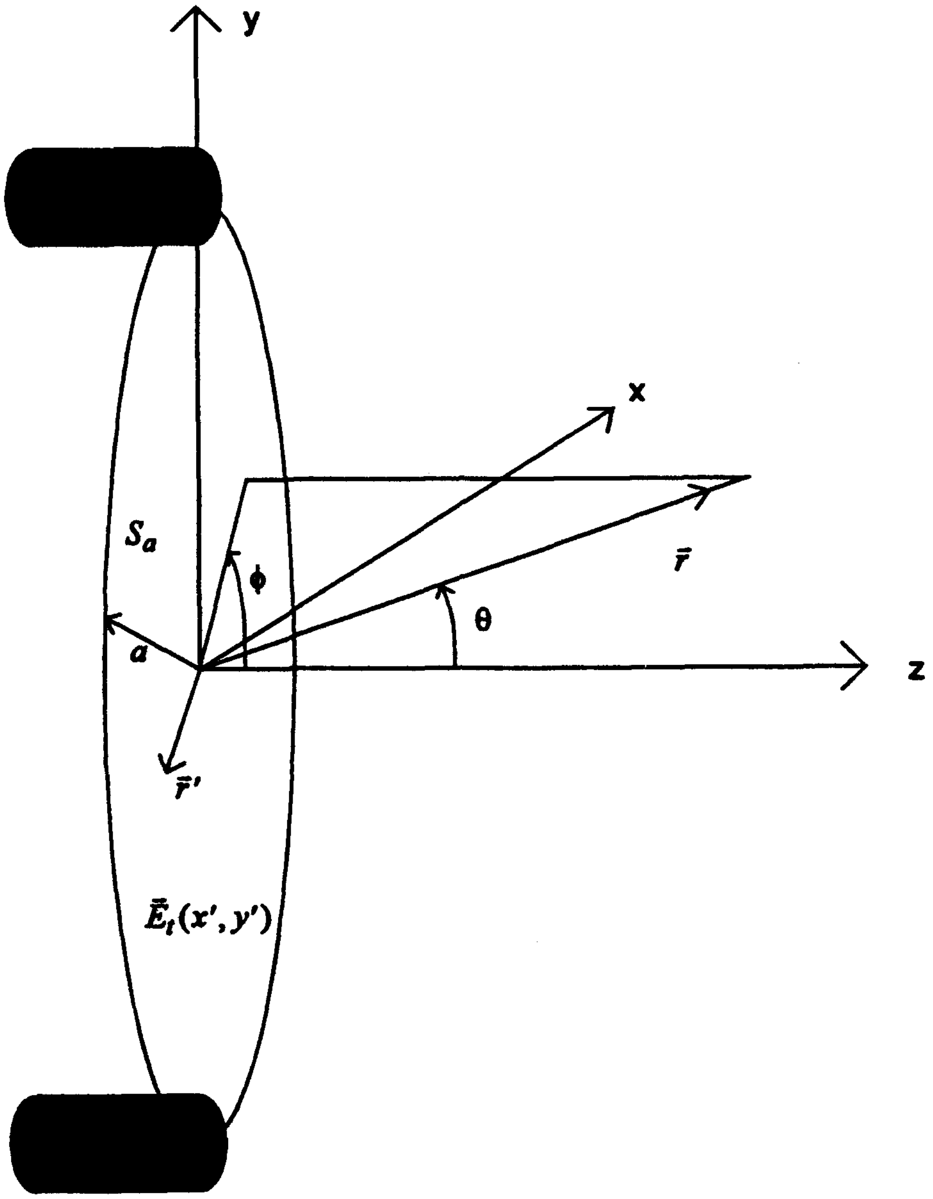

y
x
z
S_a
ϕ
θ
$\vec{r}$
a
$\vec{r}'$
$\vec{E}_t(x', y')$

Figure 2. The aperture field to be radiated.

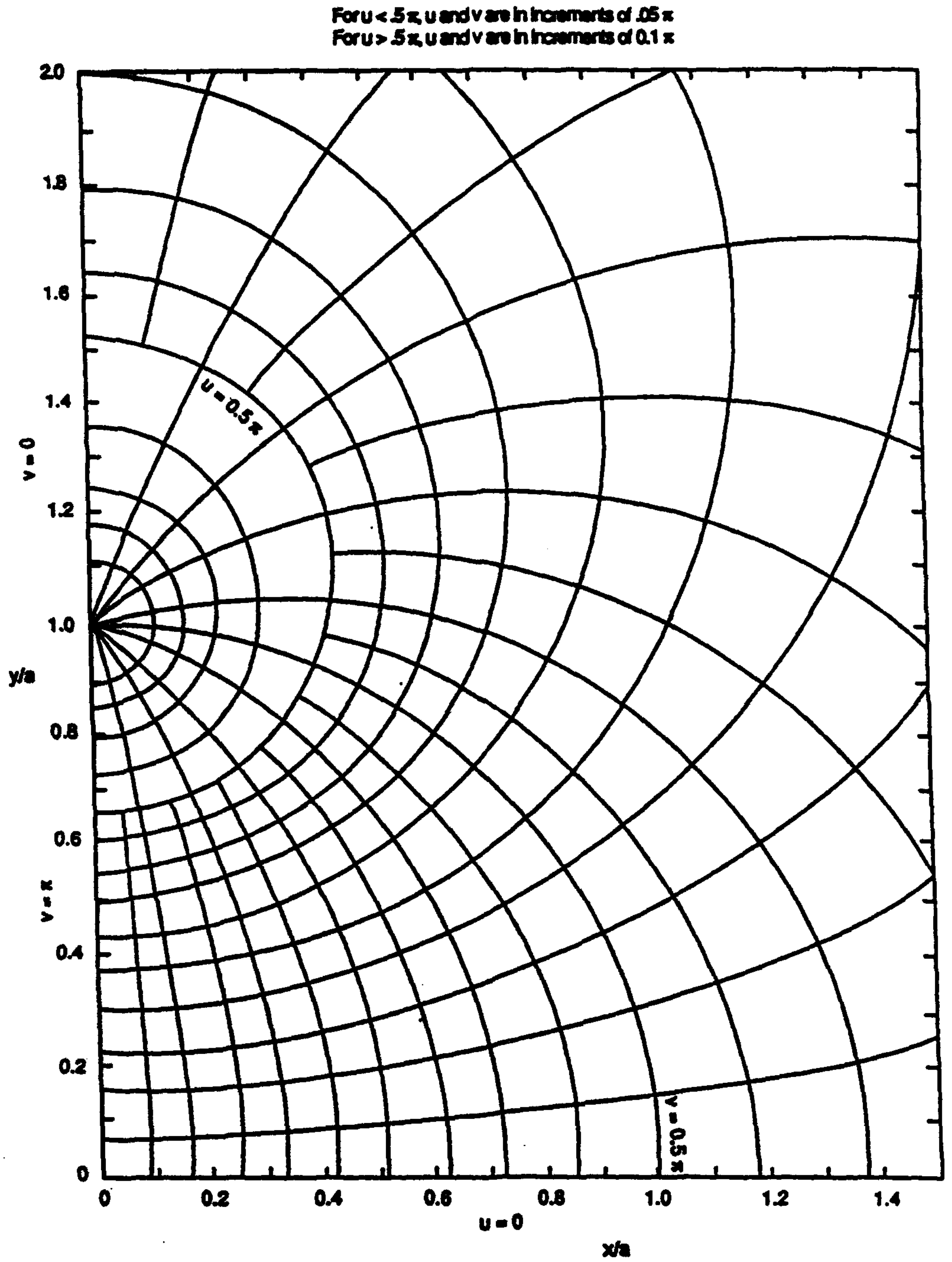


Figure 3. The complex potential map for round wires, upper right quadrant, from [2].

So far, this theory has just been a review. But in looking ahead, we know that certain additional functions will be of use to us later. Thus, let us define the following normalized potential function.

$$\Phi^{(h)}(x) \;=\; \frac{-1}{V}\int_{C_1(x)} E_y\, dy \tag{2.6}$$

where the contour $C_1(x)$ is a vertical line through the point $(x,0)$ for the length of the circular aperture (Figure 4). We will see later that the radiated field in the H-plane is proportional to the above integral. As time progresses, the field will be proportional to $\Phi^{(h)}(x)$ as the contour $C_1(x)$ sweeps across the aperture. To simplify this integral, we use a standard formula for the y component of the electric field,

$$E_y \;=\; -\frac{V}{\Delta u}\frac{\partial u(x,y)}{\partial y} \tag{2.7}$$

where $u(\zeta)$ is defined in (2.4). Thus, we have

$$\Phi^{(h)}(x) \;=\; \frac{1}{\Delta u}\int_{C_1(x)} \frac{\partial u}{\partial y}\, dy \;=\; \frac{2}{\Delta u}\, u\!\left(x, \sqrt{a^2 - x^2}\right) \tag{2.8}$$

This is just the relative change in u along the path $C_1(x)$. When C_1 cuts through the metal conductor, the relative change in u is unity. To the left and right of the conductor, the change in u tapers off, reaching zero at the edge of the aperture. Using (2.4), after some simplification we find

$$\Phi^{(h)}(x) \;=\; \begin{cases} 1/(\pi f_g)\,\mathrm{arcsech}(-x/a) & -a \le x \le -a\,\mathrm{sech}(\pi f_g) \\ 1 & -a\,\mathrm{sech}(\pi f_g) \le x \le a\,\mathrm{sech}(\pi f_g) \\ 1/(\pi f_g)\,\mathrm{arcsech}(x/a) & a\,\mathrm{sech}(\pi f_g) \le x \le a \end{cases} \tag{2.9}$$

This is plotted in Figure 5 for several values of f_g.

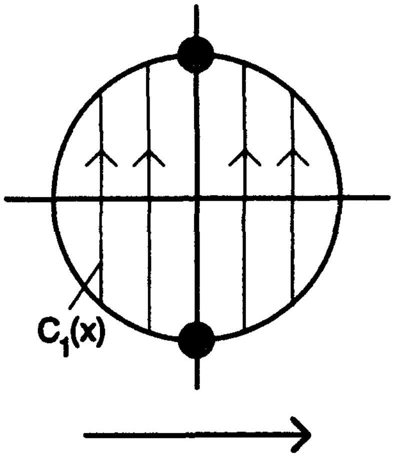


Figure 4. Location of $C_1(x)$.

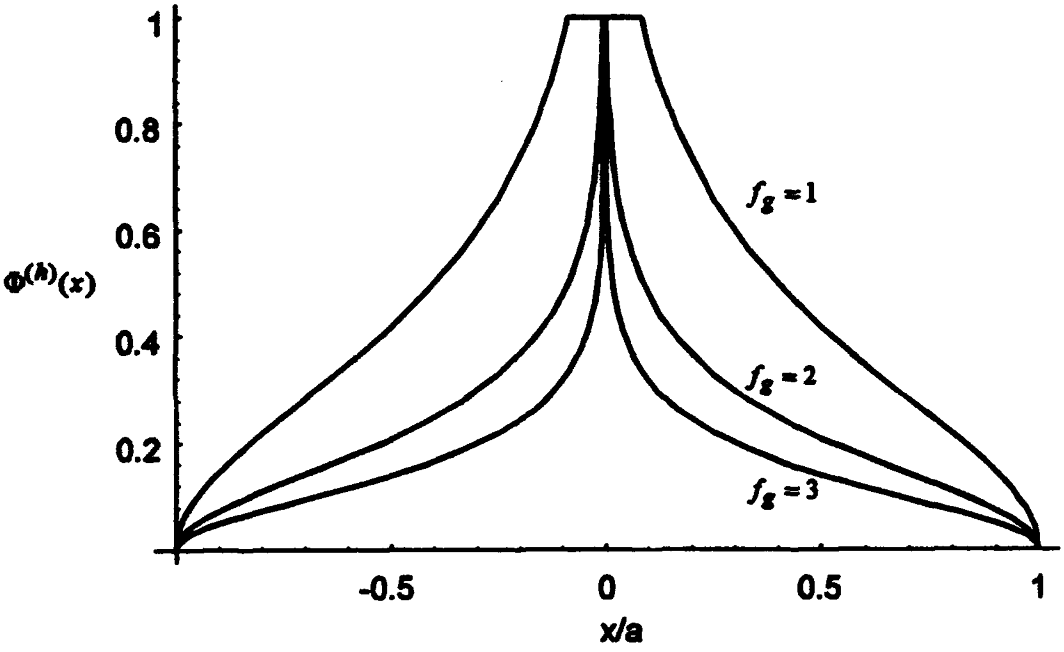


Figure 5. Plots of $\Phi^{(h)}(x)$ as a function of x/a and f_g.

There is a similar function we can define which is useful for a pattern cut in the E-plane. We will see later that the field in the E-plane is proportional to the line integral

$$\Phi^{(e)}(y) = \frac{-1}{V}\int_{C_2(y)} E_y \, dx \tag{2.10}$$

where $C_2(y)$ is a horizontal line through the point $(0, y)$, for the length of the circular aperture (Figure 6). The field in the E-plane is proportional to the above integral as $C_2(y)$ sweeps across the aperture. We now write the usual expression for E_y, and convert to a derivative with respect to x with the Cauchy-Riemann equations [9]. Thus, we have

$$E_y = -\frac{V}{\Delta u}\frac{\partial u}{\partial y} = \frac{V}{\Delta u}\frac{\partial v}{\partial x} \tag{2.11}$$

Combining the above two equations, we find

$$\Phi^{(e)}(y) = \frac{-1}{\Delta u}\int_{C_2(y)}\frac{\partial v}{\partial x}\,dx = \frac{-2}{\Delta u}\left[v\left(\sqrt{a^2-y^2}, y\right) - v(0, y)\right] \tag{2.12}$$

We can simplify the above by noticing that for almost all values of y, the expression in square brackets is $-\Delta v/4$. If $C_2(y)$ intersects one of the wires, there is a small deviation from this simple result. But at high impedances the error is negligible because the wires are small. Thus, we have

$$\Phi^{(e)}(y) = \begin{cases} 1/(2f_g) & |y| \le a \\ 0 & \text{else} \end{cases} \tag{2.13}$$

This is just a pulse function of width a and magnitude $1/(2f_g)$. Another way of expressing this is

$$\Phi^{(e)}(y) = \frac{1}{2f_g}P(y,a) \tag{2.14}$$

where $P(y,a)$ is a pulse function of value unity for $|y| \leq a$ and zero elsewhere.

This completes the review of the static fields in the aperture. Next, we apply this theory to the radiated field.

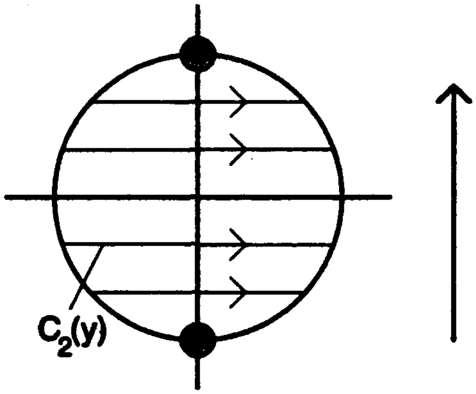


Figure 6. Location of $C_2(y)$.

3. The Radiated Field

In [1] a general expression for the radiated far field generated by an arbitrary aperture field was derived. We now wish to specialize that expression to the field resulting from the aperture field between two wires.

Let us first provide some definitions.

$$
\begin{aligned}
s &= \text{Laplace Transform variable} \\
\gamma &= s/c \\
c &= \text{speed of light} \qquad (3.1) \\
\bar{\bar{1}} &= \bar{1}_x\bar{1}_x + \bar{1}_y\bar{1}_y + \bar{1}_z\bar{1}_z \\
\bar{\bar{1}}_z &= \bar{\bar{1}} - \bar{1}_z\bar{1}_z \;=\; \bar{1}_y\bar{1}_y + \bar{1}_z\bar{1}_z
\end{aligned}
$$

Furthermore, we define the tangential aperture field as

$$
\begin{aligned}
\vec{E}_t(x',y',t) &= \left[\bar{1}_x E_x(x',y') + \bar{1}_y E_y(x',y')\right] u(t) \\
&= E_o\left[\bar{1}_x g_x(x',y') + \bar{1}_y g_y(x',y')\right] u(t) \qquad (3.2) \\
&= E_o\, \vec{g}(x',y')\, u(t)
\end{aligned}
$$

where E_o is the field at the center of the aperture. The radiated far field is now [1]

$$
\begin{aligned}
\vec{E}(\vec{r},s) &= e^{-\gamma r}\, \frac{E_o A}{2\pi c r}\, \bar{\bar{F}}(\bar{1}_r,s) \\
\bar{\bar{F}}(\bar{1}_r,s) &= \frac{1}{A}\left[(\bar{1}_z\cdot\bar{1}_r)\bar{\bar{1}} - (\bar{1}_z\bar{1}_r)\right]\cdot\int_{S_a} e^{\gamma(\bar{1}_r\cdot\vec{r}\,')}\vec{g}(x',y')\, dS'
\end{aligned}
\qquad (3.3)
$$

where we have assumed a step-function time dependence of the driving voltage. To get the result for arbitrary time dependence, one would convolve the step response with the derivative of the driving voltage.

It can be shown from symmetry arguments [7] that the radiated field in the dominant planes (E-plane and H-plane) can be represented by just the dominant polarization, without a secondary or crosspol component. Another way of saying this is that it is only E_y in the aperture that contributes to the radiated far field in the two major plane cuts. If we now specialize the above relationship to include just E_y, we have in the principal planes

$$\begin{aligned}
\tilde{\bar{E}}(\bar{r},s) &= e^{-\gamma r}\,\frac{E_0 A}{2\pi c r}\,\tilde{\bar{\bar{F}}}(\bar{1}_r,s) \\
\tilde{\bar{\bar{F}}}(\bar{1}_r,s) &= \left[\bar{1}_r \times \bar{1}_x\right] \tilde{f}_a(\bar{1}_r,s) \\
&= \left[\sin(\phi)\bar{1}_\theta + \cos(\theta)\cos(\phi)\bar{1}_\phi\right] \tilde{f}_a(\bar{1}_r,s) \\
\tilde{f}_a(\bar{1}_r,s) &= \frac{1}{A}\int_{S_a} e^{\gamma(\bar{1}_r\cdot\bar{r}')} g_y(x',y')\,dS'
\end{aligned} \tag{3.4}$$

This is the expression we integrate over to get the field for the dominant polarization. Note that we have used the vector expression

$$\bar{A}\times(\bar{B}\times\bar{C}) = (\bar{A}\cdot\bar{C})\bar{B} - (\bar{A}\cdot\bar{B})\bar{C} \tag{3.5}$$

to get the result in (3.4).

Let us simplify further (3.4), which is the complete radiated far field in one of the principal planes. The y component of the aperture electric field can be expressed in various ways as

$$E_{t_y}(x',y') = E_0\, g_y(x',y') = -\frac{V}{\Delta u}\frac{\partial u}{\partial y} \tag{3.7}$$

so the function g_y is

$$g_y(x',y') = -\frac{V}{E_0\,\Delta u}\frac{\partial u(x,y)}{\partial y} \tag{3.8}$$

The function $\tilde{f}_a(\bar{1}_r,s)$ in (3.4) now simplifies to

$$\tilde{f}_a(\bar{1}_r,s) = -\frac{V}{A E_0\,\Delta u}\int_{S_a} e^{\gamma(\bar{1}_r\cdot\bar{r}')}\,\frac{\partial u(x',y')}{\partial y'}\,dS' \tag{3.9}$$

Converting to the time domain, we find

$$f_a(\bar{1}_r,t) = -\frac{V}{AE_o\,\Delta u}\int_{S_a}\frac{\partial u(x',y')}{\partial y'}\,\delta(t+\bar{1}_r\cdot\bar{r}'/c)\,dS' \qquad (3.10)$$

This is the result we must evaluate in both the E- and H-planes.

In the H-plane, where $\phi = 0$ or π, the function $f_a(\bar{1}_r,t)$ simplifies to

$$\begin{aligned} f_a^{(h)}(\theta,t) &= -\frac{V}{AE_o\,\Delta u}\frac{c}{\sin(\theta)}\int_{C_a(x)}\frac{\partial u(ct/\sin(\theta),y')}{\partial y'}\,dy' \\ &= -\frac{V}{AE_o}\frac{c}{\sin(\theta)}\,\Phi^{(h)}(ct/\sin(\theta)) \end{aligned} \qquad (3.11)$$

where we have used the function $\Phi^{(h)}(x)$ as defined in Section II. The superscript *h* indicates that the expression is specialized to the H-plane. Combining this with (3.4), and noting that $\phi = 0$ or π in the H-plane, we find

$$\boxed{\bar{E}^{(h)}(r,t) = \bar{1}_y\left(\frac{-V}{r}\right)\frac{\cot(\theta)}{2\pi}\,\Phi^{(h)}\!\left(\frac{ct}{\sin(\theta)}\right)} \qquad (3.12)$$

Recall that in the H-plane, $\bar{1}_y = \pm\,\bar{1}_\phi$, where the "+" sign applies if $\phi = 0$ and the "-" sign applies for $\phi = \pi$. This is the final expression we need for the field in the H-plane. Note that we have normalized out a time delay of *r/c*, to keep the expression simple.

We can now plot some examples of these fields (Figure 7). If we normalize the time scale properly, the step response is dependent solely upon f_g and θ. Thus, it is convenient to normalize the time scale to $t_a = a/c$.

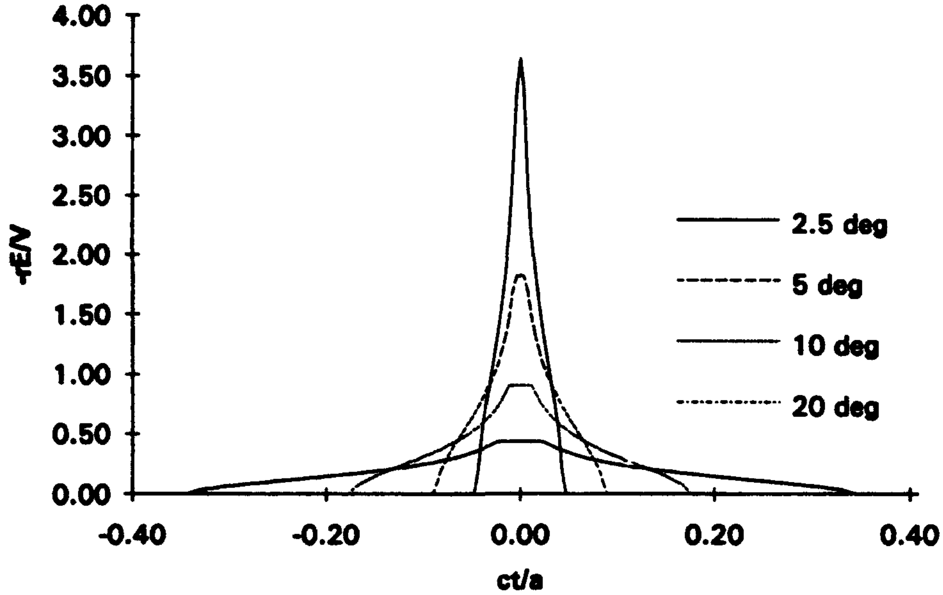


Figure 7. H-plane step response, for various values of θ. For these plots, $Z_c = 400\ \Omega$ ($f_g = 1.0631$).

We can find similar expressions now for the E-plane. Proceeding from (3.10), and confining ourselves to the E-plane, where $\phi = \pi/2$ or $3\pi/2$, we find

$$\begin{aligned} f_a^{(e)}(\theta,t) &= -\frac{V}{AE_o\,\Delta u}\frac{c}{\sin(\theta)}\int_{C_2(y)}\frac{\partial u(x',ct/\sin(\theta))}{\partial y'}dx' \\ &= \frac{V}{AE_o\,\Delta u}\frac{c}{\sin(\theta)}\int_{C_2(y)}\frac{\partial v(x',ct/\sin(\theta))}{\partial x'}dx' \end{aligned} \tag{3.13}$$

Expressing this in terms of the normalized potential function for the E-plane, $\Phi^{(e)}(y)$, we find

$$\begin{aligned} f_a^{(e)}(\theta,t) &= -\frac{V}{AE_o}\frac{c}{\sin(\theta)}\Phi^{(e)}(ct/\sin(\theta)) \\ &= -\frac{V}{AE_o}\frac{c}{\sin(\theta)}\frac{1}{2f_g}P(x,a) \end{aligned} \tag{3.14}$$

where $P(x,a)$ is the pulse function described in Section 2 of this paper. Combining this with (3.4), and noting that $\phi = \pi/2$ or $3\pi/2$ in the E-plane, we have

$$\boxed{\begin{aligned}\vec{E}^{(e)}(r,\theta,t) &= \pm\vec{1}_\theta\left(\frac{-V}{r}\right)\frac{1}{2\pi\sin(\theta)}\;\Phi^{(e)}\left(\frac{ct}{\sin(\theta)}\right)\\ &= \pm\vec{1}_\theta\left(\frac{-V}{r}\right)\frac{1}{4\pi f_g\sin(\theta)}\;P\left(\frac{ct}{\sin(\theta)},a\right)\end{aligned}} \tag{3.15}$$

where the "+" applies if $\phi = \pi/2$, and the "-" sign applies if $\phi = 3\pi/2$. This is the final expression for the field in the E-plane due to the y component of the aperture field. Some examples of these fields are plotted in Figure 8.

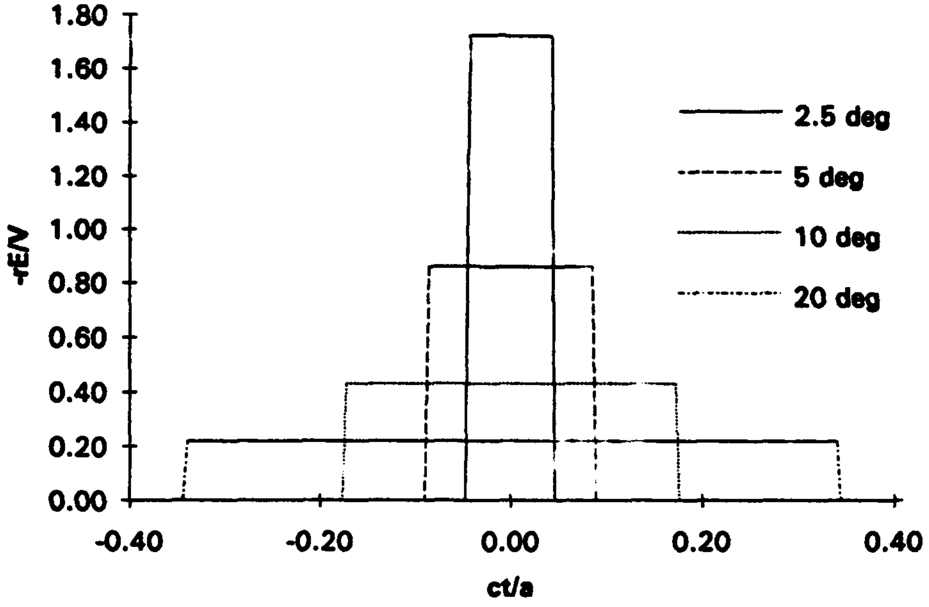


Figure 8. E-plane step response, for various values of θ. For these plots, $Z_c = 400\ \Omega$ ($f_g = 1.0631$).

Let us review now the approximations we have used. We have ignored aperture blockage, which limits us to higher feed impedances. We have also ignored some of the fine detail in the E-plane step response that occurs close to the wires. This has the effect of making the E-plane step response a slightly sharper pulse function than it actually is. We have not invoked any small-angle approximations, however, so our expressions should be approximately valid out to $\theta = 90^{\circ}$.

We now have the step response in the E- and H-planes of an IRA in equations (3.12) and (3.15). Before going further, we check that our two expressions provide the correct result on boresight.

4. Consistency Check on Boresight

We need to check that the expressions we have just developed for the H-plane and E-plane fields generate the correct results on boresight. Recall that the high-impedance approximation on boresight calculated by contour integration is [2]

$$E(r,t) = \frac{-V}{r} \frac{a}{2\pi c f_g} \delta_a(t) \tag{4.1}$$

where $\delta_a(t)$ is the approximate delta function [1], which converges to a true delta function in the limit as $r \to \infty$. Thus, the expressions for the E-plane and H-plane fields must approach this answer at $\theta = 0$, and at high impedances. Since it is simplest to show this for the E-plane result, we do this first.

The E-plane expression for the field on boresight is

$$\bar{E}^{(e)}(r,\theta,t) = \pm \bar{1}_{\theta} \left(\frac{-V}{r}\right) \frac{1}{4\pi f_g \sin(\theta)} P\left(\frac{ct}{\sin(\theta)}, a\right) \tag{4.2}$$

In comparing (4.1) and (4.2), we find that if the two expressions are equivalent on boresight, then the following must also be true

$$\delta(t) = \lim_{\theta\to 0} \frac{c}{2a\sin(\theta)} P\left(\frac{ct}{\sin(\theta)}, a\right) \tag{4.3}$$

Recall now the characteristics of a function that make it a δ function. First, the value of the function must approach infinity at $t = 0$. For the above function, this is obviously true. Second, the integral of the function (area) must be unity. Integrating (4.3), we find

$$\begin{aligned} \lim_{\theta\to 0} \frac{c}{2a\sin(\theta)} \int P\left(\frac{ct}{\sin(\theta)}, a\right) dt &= \lim_{\theta\to 0} \frac{c}{2a\sin(\theta)} \int_{-a}^{a} \frac{\sin(\theta)}{c} P(y,a)\, dy \\ &= 1 \end{aligned} \tag{4.4}$$

Thus, the area under the curve is unity, and the function approaches the δ function on boresight. Therefore, our E-plane expression is consistent with previous results on boresight.

In the H-plane, the proof is only a little more complicated. The field we calculated is

$$\vec{E}^{(h)}(r,t) = \vec{1}_y \left(\frac{-V}{r}\right) \frac{\cot(\theta)}{2\pi}\, \Phi^{(h)}\left(\frac{ct}{\sin(\theta)}\right) \tag{4.5}$$

Comparing this with (4.1), we find that for the two to be the same on boresight, it must also be true that

$$\delta_a(t) = \lim_{\theta\to 0} \frac{c\, f_g \cot(\theta)}{a}\, \Phi^{(h)}\left(\frac{ct}{\sin(\theta)}\right) \tag{4.6}$$

where

$$\Phi^{(h)}(x) \;=\; \begin{cases} 1/(\pi f_g)\,\mathrm{arcsech}(-x/a) & -a \le x \le -a\,\mathrm{sech}(\pi f_g) \\ 1 & -a\,\mathrm{sech}(\pi f_g) \le x \le \;a\,\mathrm{sech}(\pi f_g) \\ 1/(\pi f_g)\,\mathrm{arcsech}(x/a) & a\,\mathrm{sech}(\pi f_g) \le x \le \;a \end{cases} \tag{4.7}$$

Clearly the expression in (4.6) approaches infinity at $t = 0$. Next we check that the integral is equal to unity. Integrating (4.6), we find

$$\begin{aligned} \lim_{\theta\to 0} \frac{c\,f_g \cot(\theta)}{a} \int \Phi^{(h)}\!\left(\frac{ct}{\sin(\theta)}\right) dt \;&=\; \lim_{\theta\to 0} \frac{c\,f_g \cot(\theta)}{a} \int_{-a}^{a} \frac{\sin(\theta)}{c}\,\Phi^{(h)}(x)\,dx \\ &=\; \frac{f_g}{a} \int_{-a}^{a} \Phi^{(h)}(x)\,dx \end{aligned} \tag{4.8}$$

To carry out the integral, we need the result from Dwight[8, #738], that

$$\int \mathrm{arcsech}(x/a)\,dx \;=\; x\,\mathrm{arcsech}(x/a) \;+\; a\,\arcsin(x/a) \qquad \left[\mathrm{arcsech}(x/a) \;>\; 0\right] \tag{4.9}$$

This leads to a final result for the integral of

$$\begin{aligned} \int_{-a}^{a} \Phi^{(h)}(x)\,dx \;&=\; 2a\,\mathrm{sech}(\pi f_g) \;+\; \frac{2}{\pi f_g} \int_{a\,\mathrm{sech}(\pi f_g)}^{a} \mathrm{arcsech}(x/a)\,dx \\ &=\; \frac{a}{f_g}\left(1 - \frac{2}{\pi}\arcsin(\mathrm{sech}(\pi f_g))\right) \end{aligned} \tag{4.10}$$

At high impedances $\mathrm{sech}(\pi f_g) \to 0$, so the second term in the above expression approaches 0. Combining this with (4.8), we find the area under the function is

$$\lim_{\theta \to 0} \frac{c\, f_g \cot(\theta)}{a} \int \Phi^{(h)}\left(\frac{ct}{\sin(\theta)}\right) dt \;=\; 1 \tag{4.11}$$

This is the second condition required for the H-plane field to approach a δ function on boresight. Thus, the proof is complete and all our off-boresight fields are consistent with known results on boresight.

5. Sample Field Calculations

Now that we are assured that the step responses are correct, we use them to plot some sample fields. To do so, we convolve the derivative of a typical driving voltage with the step response.

We use the integral of a Gaussian waveform to drive the antenna. Thus, the step response is convolved with a Gaussian waveform to get the actual radiated field. It is convenient to define the driving waveforms as

$$\frac{dv(t)}{dt} = \frac{V}{t_d} e^{-\pi(t/t_d)^2}, \qquad t_{FWHM} = 0.940\, t_d \tag{5.1}$$

$$v(t) = \int_{-\infty}^{t} \frac{dv(t')}{dt'}\, dt', \qquad t_{10-90} = 1.023\, t_d \tag{5.2}$$

where t_{FWHM} is the Full Width Half Max of dv/dt, and $t_{10\text{-}90}$ is the 10-90% risetime of $v(t)$. These waveforms are plotted in Figure 9. Recall the definition of the derivative risetime of a waveform is

$$t_d = \frac{\max\,(v(t))}{\max\left(dv(t)\,/\,dt\right)} \tag{5.3}$$

Thus, the derivative risetime is inversely proportional to the maximum derivative of the driving voltage. This is a useful property, since the peak of the radiated field is proportional to the peak of the derivative of the voltage. Note also that for the integrated Gaussian waveform shown above, the derivative risetime t_d is within two percent of the 10-90% risetime, $t_{10\text{-}90}$.

The above expression for $v(t)$ can also be expressed in terms of the complementary error function. Thus,

$$v(t) = \begin{cases} V\left[\quad (1/2)\,\mathrm{erfc}(\sqrt{\pi}\,|t|/t_d)\right] & t < 0 \\ V\left[1 - (1/2)\,\mathrm{erfc}(\sqrt{\pi}\,t/t_d)\right] & t \geq 0 \end{cases} \tag{5.4}$$

where the complementary error function is defined in [11].

The radiated field is now calculated simply from

$$E(r,\theta,\phi,t) = \frac{1}{V}\frac{dv(t)}{dt} \circ E^{step}(r,\theta,\phi,t) \tag{5.5}$$

where the $\circ$ operator indicates a convolution, and $E^{step}(r,\theta,\phi,t)$ is the step response in the E- or H-plane, as calculated in Section III of this paper. We can reduce the number of cases that need to be calculated by defining a rise parameter

$$T_d = \frac{t_d}{t_a} = \frac{c\,t_d}{a} \tag{5.6}$$

All problems with equal rise parameters have the same shape radiated field. Thus, a field pattern is a function only of T_d, f_g, and θ. There is no need to include a dependence upon both the aperture radius a and the derivative risetime, t_d, because all the information is contained in T_d.

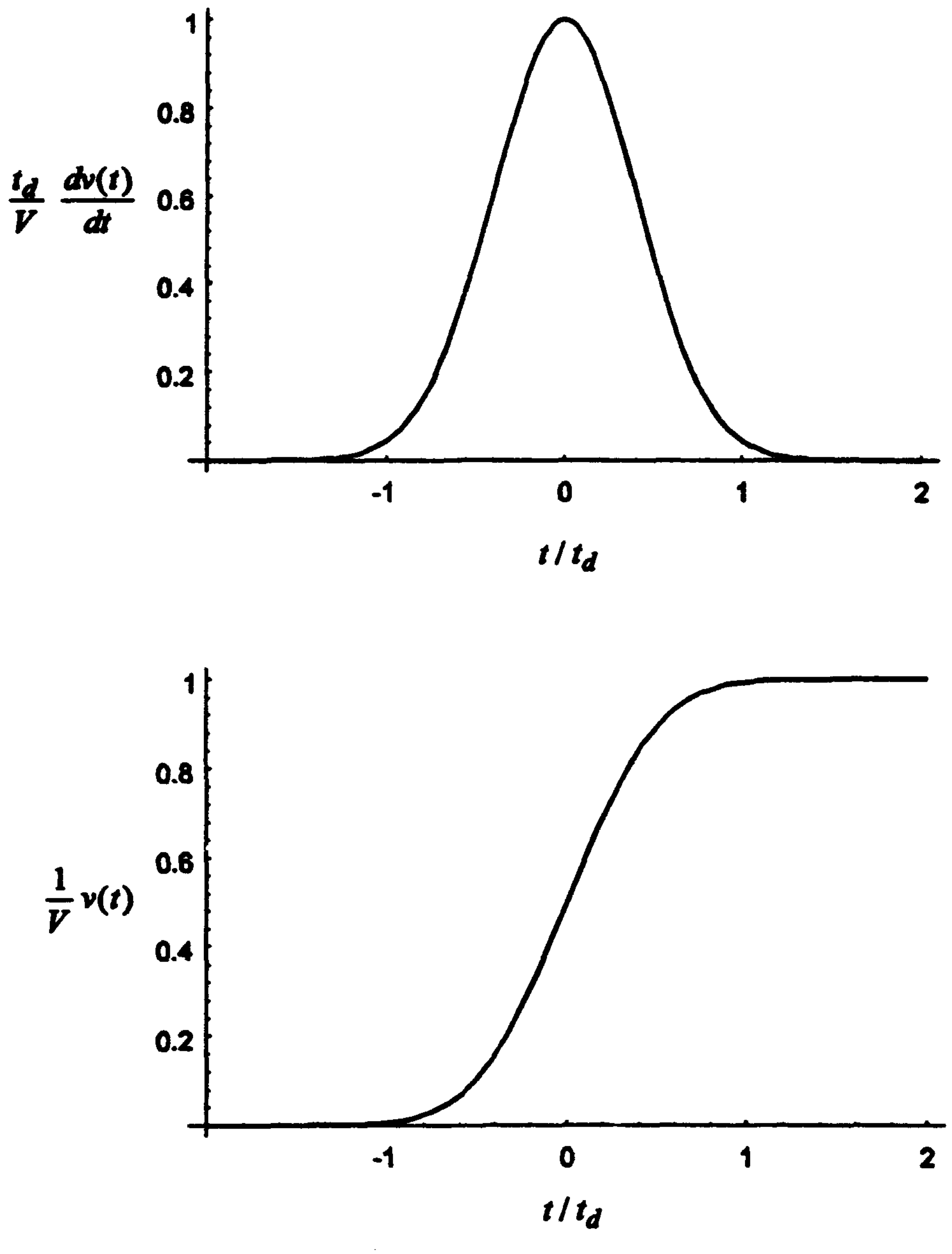


Figure 9. Derivative of driving voltage (top), and its integral (bottom).

Let us now provide a typical example. Consider an IRA with a 400 Ω feed impedance and a radius of 0.3 m, driven by an integrated Gaussian voltage with $t_d = 250$ ps ($t_{10\text{-}90} = 256$ ns). For this configuration, $T_d = 0.25$, so all possible configurations with $T_d = 0.25$ and $Z_c = 400\ \Omega$ can be plotted on the same graph, with proper scaling. The fields in the E-plane and H-plane are shown in Figure 10. The time scale has to be normalized to either t_a or t_d ; it does not matter which. We have chosen to normalize to $t_a = a/c$,. For this configuration $t_a = 1$ ns, so it is trivial to normalize the x axis.

With these results, we may now establish an antenna pattern. We do so in the section that follows.

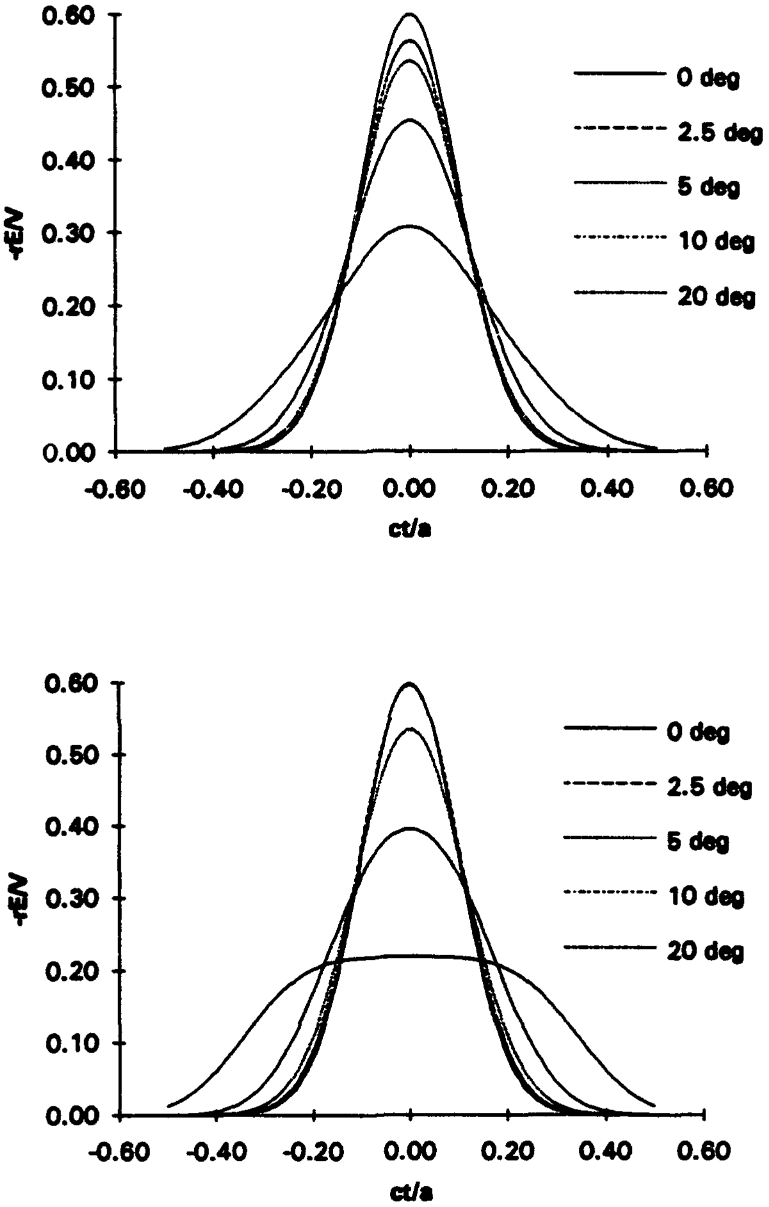


Figure 10. Radiated electric field vs. θ and t in the H-plane (top) and E-plane (bottom) for $Z_c = 400\ \Omega$ and $T_d = 0.25$. Note that the curves in the E-plane at 0° and 2.5° nearly overlap.

6. Antenna Pattern

With the off-boresight fields now calculated (at least approximately), we can establish the antenna pattern. Recall from [4] the gain of an antenna in a particular direction for transmit mode is

$$G(\theta,\phi) \;=\; \frac{2\pi c\sqrt{f_g}\left\| rE^{rad}(r,\theta,\phi,t)\right\|}{\left\| dV^{inc}(t)/dt\right\|} \tag{6.1}$$

It is simplest to think of the norm operator $\| \; \|$ as being the peak of the waveform (∞–norm). Of course, other interpretations are possible, such as the area under the curve (1-norm) or square root of power in the waveform (2-norm). Recall the definition of an arbitrary p-norm

$$\| f(t)\|_p \;\equiv\; \left(\int_{-\infty}^{\infty}|f(t)|^p\,dt\right)^{1/p}, \qquad \| f(t)\|_\infty \;\equiv\; \sup_t |f(t)| \tag{6.2}$$

The norm one chooses must correspond to the experimental technique used in detecting the pulse. Thus, when we use the peak norm (∞–norm) in these calculations, we assume an experimental system that responds to the peak in the received waveform, as opposed to the power in the waveform (2-norm) or area in the waveform (1-norm).

The above gain definition may seem unusual at first, because it defines a gain with units of meters. This definition, however, has a critical property that other definitions sometimes lack; i.e., it provides information about antenna performance in receive mode. Thus, in receive mode, if the incident waveshape is a derivative of the driving voltage used in transmit mode (in this case a Gaussian), then the received voltage wave is described by

$$\left|V^{rec}(t,\theta,\phi)\right| = \sqrt{f_g}\; G(\theta,\phi)\left|E^{inc}(\theta,\phi,t)\right| \tag{6.3}$$

In other words, the peak received voltage is a simple function of the peak incident field. Some other gain definitions do not offer such simple interpretations consistent with reciprocity.

We now plot the gain as a function of angle for the configuration of the previous section (Figure 11). It is interesting to note that on boresight, the gain is just the aperture radius divided by $f_g^{1/2}$. Furthermore, the pattern in the E-plane is somewhat more compressed than that in the H-plane.

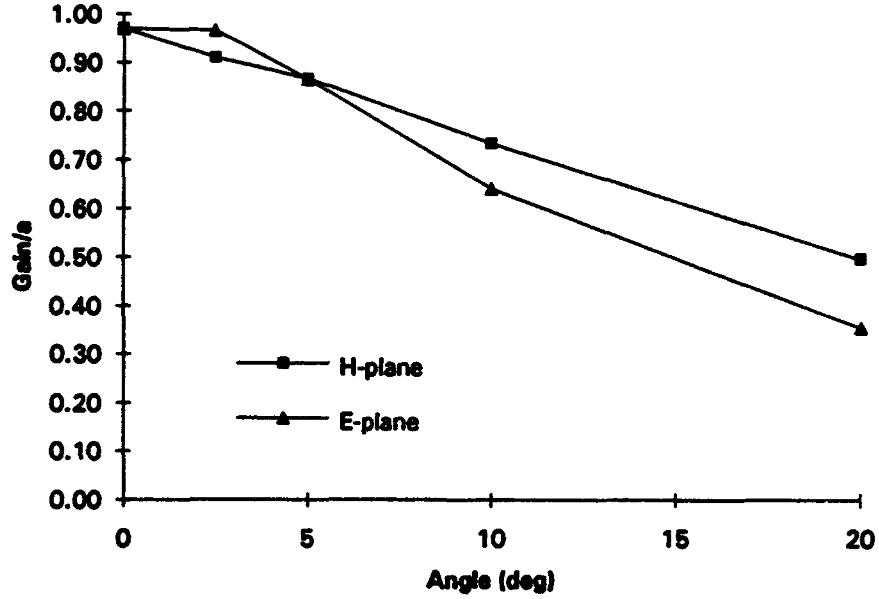


Figure 11. Gain (∞-norm) in the H- and E-planes as a function of angle. For this plot, $Z_c = 400\ \Omega$, and $T_d = 0.25$ ($a = 0.3$ m and $t_d = 0.25$ ns).

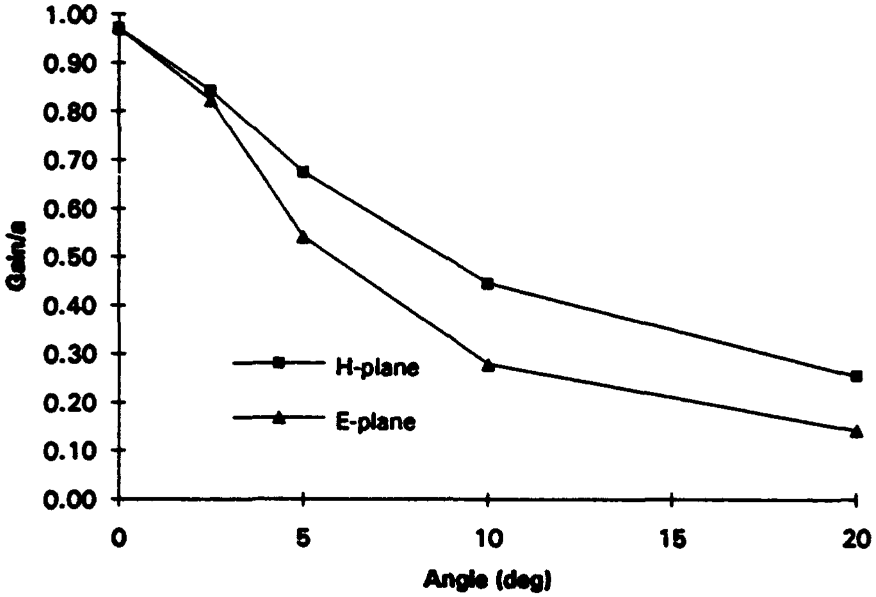


Figure 12. Gain (∞-norm) in the H- and E-planes as a function of angle. For this plot, $Z_c = 400\ \Omega$, and $T_d = 0.1$ ($a = 0.3$ m and $t_d = 0.1$ ns).

The particular choice of parameters we have used until now has a slower rise parameter T_d than what may be more typical. In order to see the difference, we plot the gain for a faster configuration, with $T_d = 0.1$ and the same feed impedance, $Z_c = 400\ \Omega$. If we assume the same radius as before ($a = 0.3$ m), we have $t_d = 0.1$ ns., which is 2.5 times faster than the first case. The gain for this arrangement is shown in Figure 12. The overall effect is to make the patterns narrower. Also, we can see here more clearly that the E-plane pattern is narrower than the H-plane pattern.

It may be surprising, at first, to notice that the boresight gains of the two arrangements are the same. The faster configuration has a larger radiated field on boresight by a factor of 2.5, so one might think that the boresight gain must also be 2.5 times larger. But our definition of gain in (6.1) is such that the faster risetime appears in the denominator of (6.1) (the dv/dt term), not in the gain. One could define a gain that included the factor of 2.5, but it's meaning in receive mode would then become less clear.

We can define a beamwidth as being the angular width where the pattern is down by a factor of two from its peak. Since this angular width depends upon the selection of the norm used in the gain definition, we call this beamwidth the Half Norm Beam Width (HNBW), analogous to the full-width half max (FWHM) beamwidth commonly used in the frequency domain. At higher T_d's (larger apertures or faster risetimes) the beamwidth becomes narrower.

7. Discussion

There are a number of ways of controlling antenna pattern. Let us assume we have calculated a pattern, and we wish to make it broader. There are several ways of doing so.

The first technique for broadening pattern is to defocus the feed point of the aperture. Although this is conceptually simple, no simple models exist yet for predicting the effect. If one wanted to pursue this, one might use a technique similar to that used to predict TEM horn performance [12], which is another defocused aperture.

The second technique of broadening an antenna pattern is to use a smaller radius aperture. The third technique is to use a slower driving waveform (with a larger t_d). Both of these last two techniques increase the rise parameter T_d. In doing so, the on-boresight radiated field is reduced in inverse proportion to T_d. If one defocuses the antenna, the on-boresight radiated field is similarly reduced, but it is somewhat more difficult to predict the magnitude of this reduction.

The final technique for broadening an antenna pattern is to change the norm one is using to detect the signal. Recall that the norm one chooses in the gain definition must correspond to the experimental method of detecting the pulse. We have used the peak norm (∞-norm) in this paper, however, both the 2-norm (power) and 1-norm (area) are reasonable choices. Furthermore, it can be shown that the 2-norm provides a broader pattern than the ∞-norm, and the 1-norm is broader still. Thus, if one is detecting a pulse using peak detection (∞-norm), one could get a broader antenna pattern by detecting the power in the pulse (2-norm). One would get a broader pattern still by detecting the integral over the pulse (1-norm).

Conversely, if one wanted a narrower beam with a larger signal on boresight, one could use the opposite of these techniques. Thus, one would use large, well-focused apertures with fast risetimes (small T_d). Furthermore, one would detect the signal using peak detection (∞-norm).

8. Conclusion

We have demonstrated a simple technique for calculating the early-time off-boresight radiated field generated by a reflector Impulse Radiating Antenna. Aperture blockage is not included in the model, so the model is valid only at higher feed impedances. This model leads to a simple interpretation of transient antenna gain, which we plot for some sample cases. Because the models we developed are simple, one can readily understand how to control the antenna pattern by adjusting antenna parameters.

Acknowledgment

We wish to thank Mr. John McCorkle of the Army Research Laboratory and Mr. Dale A. Steffen of Electro Magnetic Applications, Inc. for funding portions of this work.

References

1. Baum, C. E. (1989) Radiation of Impulse-Like Transient Fields, Sensor and Simulation Note 321.

2. Baum, C. E. (1991) Aperture Efficiencies for IRAs, Sensor and Simulation Note 328.

3. Farr, E. G. (1993) Optimizing the Feed Impedance of Impulse Radiating Antennas, Part I: Reflector IRAs, Sensor and Simulation Note 354.

4. Farr, E. G. and Baum, C. E. (1992) Extending the Definitions of Antenna Gain and Radiation Pattern Into the Time Domain, Sensor and Simulation Note 350.

5. Farr, E. G. and Baum, C. E. (1992) Prepulse Associated with the TEM Feed of an Impulse Radiating Antenna, Sensor and Simulation Note 337.

6. Smythe, W. R. (1989) *Static and Dynamic Electricity*, third ed., New York: Hemisphere.

7. Baum, C. E. and Kritikos, H. N. (1990) Symmetry in Electromagnetics, Physics Note 2.

8. Dwight, H. B. (1961) *Tables of Integrals and Other Mathematical Data*, fourth ed., New York: Macmillan Publishing Co.

9. Churchill, R. V., et al (1976) *Complex Variables and Applications*, third ed., New York: McGraw-Hill.

10. Baum, C. E. (1966) Impedances and Field Distributions for Symmetrical Two Wire and Four Wire Transmission Line Simulators, Sensor and Simulation Note 27.

11. Abramowitz, M. and Stegun, I. A. (1972) *Handbook of Mathematical Functions*, National Bureau of Standards.

12. Farr, E. G. and Baum, C. E. (1992) A Simple Model of Small-Angle TEM Horns, Sensor and Simulation Note 340.

THE EFFICIENCY OF ORTHOGONAL TRANSFORM DOMAIN ADAPTIVE PROCESSORS FOR WIDEBAND MULTICHANNEL ARRAYS

J. S. Goldstein,[1,2] M. A. Ingram,[2] P. D. Anderson,[2] and A. D. Forrest[2]

[1] USAF Rome Laboratory, Space Communications Branch
525 Brooks Rd., Griffiss AFB, NY 13441-4505
and
[2] School of Electrical Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250

ABSTRACT: The transformation of array observation data prior to adaptive sidelobe cancellation with the LMS algorithm may provide an improvement in the adaptive processor performance. The use of a time-varying step size is a necessary, but not sufficient, condition for any improvement compared to the standard LMS algorithm. The ability of the transform to decorrelate the observation data determines the efficiency of the time-varying step size LMS processor. The optimal Karhunen-Loève transform and sub-optimal approximations are considered. The eigenvalue spread and mean-square error performance measures of the transform domain processor are considered graphically as a function of jammer spatial arrival angle and temporal frequency. Assuming stationarity, these three-dimensional plots are provided for the identity operator, DFT and DCT time-invariant transforms in a one desired signal - one jammer scenario.

1 Introduction

This paper addresses the problem of improving the convergence speed of the least mean squares (LMS) algorithm. The application considered is linearly constrained wideband sensor array processing with a Generalized Sidelobe Canceler (GSC) form processor.

The LMS algorithm [1] is a member of the stochastic gradient class of algorithms. It is a method to descend down the mean-square error contour in the direction of the negative gradient. Its form is

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu e(k)\mathbf{X}(k), \tag{1}$$

where μ is the step size of the iterative descent, $e(k)$ is the error signal, given by the difference between a desired signal $d(k)$ and the array output $y(k)$: $e(k) = d(k) - y(k)$, and $\mathbf{X}(k)$ is the observation data. The term $e(k)\mathbf{X}(k)$ represents the instantaeous gradient estimate for the weight update.

Another popular class of algorithms used in adaptive arrays, the least squares class, includes Recursive Least Squares (RLS) [2] and Sample Matrix Inversion (SMI) [3]. Stochastic gradient algorithms are characterized by computational simplicity and an undesireable dependency between their speed of convergence and the signal environment in which they operate. Least squares algorithms are more complex and cost more in terms of computational requirements, but exhibit a relative invariance to the signal environment. Because of physical constraints, such as the platform's size, weight, and power restrictions, the processor may be subject to limitations on computational complexity. Thus the simple LMS algorithm is attractive for use in the adaptation of the array weights. The application often requires fast performance, such as in the case of anti-jam protection for satellite communicatons. These conflicting goals of simplicity and speed motivate the need to find a methodology for increasing the speed of convergence of LMS while maintaining its computational simplicity.

The paper presents a clear relationship between the stochastic gradient and least squares classes and demonstrates that it is possible to achieve the performance benefits of least squares while maintaining the computational requirements of LMS. The use of a time-varying step size coupled with an orthogonal transform is the methodology applied to achieve this goal. The time-varying step size is a necessary, but not sufficient, condition, and the efficiency of the orthogonal transform in terms of its energy decorrelating and compacting properties determines the degree to which a performance increase can be realized.

2 Linearly Constrained Adaptive Array Processing

Consider the K-sensor, J-tap uniform linear antenna array shown in Figure 1. The desired signal $\mathbf{S}(k)$ and noise $\mathbf{N}(k)$ define the received data vector at the kth sample as $\mathbf{X}(k) = \mathbf{S}(k) + \mathbf{N}(k)$. The signal $\mathbf{X}(k)$ impinges the sensor array and arrives at each element at a different time determined by the array spacing and the composite signal's direction of arrival. The K-dimensional vector of the values of the received waveform $\mathbf{X}(k)$ prior to the

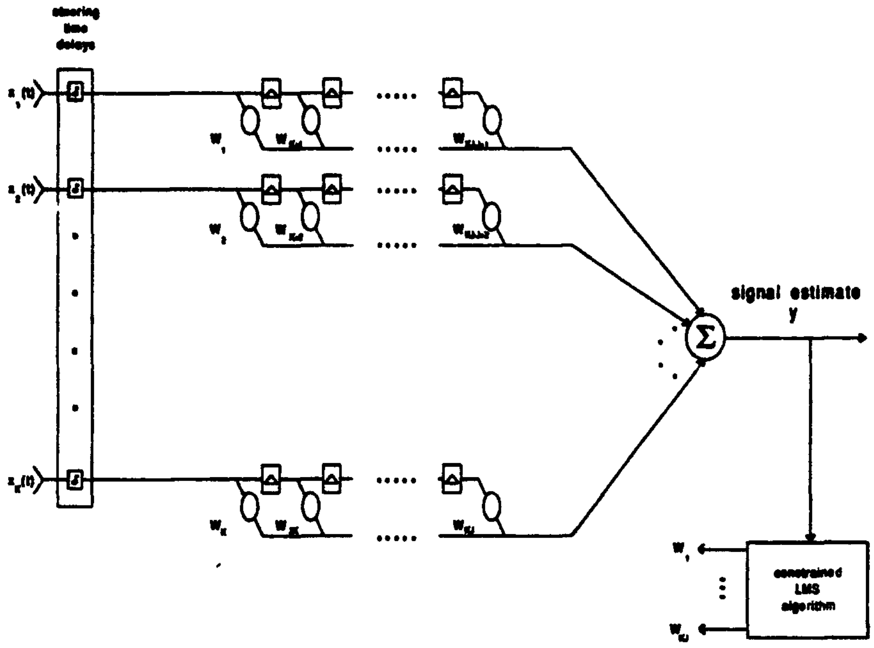


Figure 1: A direct form constrained processor.

i^{th} tap of the tapped delay lines is written as

$$\mathbf{X}_i(k) = \begin{bmatrix} x_i(t) \\ x_i(t-\tau) \\ \cdot \\ \cdot \\ x_i(t-(K-1)\tau) \end{bmatrix}, \qquad i \; = \; 1, 2, \ldots J, \tag{2}$$

and the resulting KJ-dimensional stacked data vector is formed as

$$\mathbf{X}(k) = \begin{bmatrix} \mathbf{X}_1(k) \\ \mathbf{X}_2(k) \\ \cdot \\ \cdot \\ \mathbf{X}_J(k) \end{bmatrix}. \tag{3}$$

The sensor elements are equispaced by a distance d, and $\tau = d \sin(\theta)/c$ is the interelement delay for a plane wave impinging the elements of the array from an angle θ with speed c.

Griffiths and Jim [4] have shown that the linearly constrained array introduced by Frost [5] is equivalent to the GSC in Figure 2, where the constraint is imposed via the full-rank $M \times K$ matrix filter $\mathbf{W}_S$ and the conventional beamformer $\mathbf{W}_C$. The GSC reduces the dimension of the observation data by the number of constraints, and for the single constraint

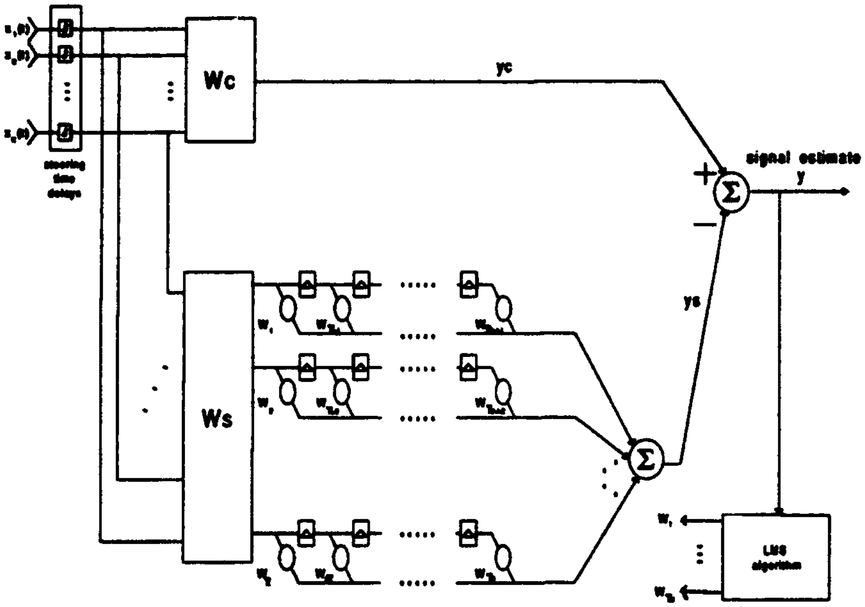


Figure 2: Tapped-delay-line form of the GSC.

case of interest, $M = \tilde{K} = K - 1$. The signal blocking matrix filter must satisfy

$$\mathbf{r}_i^T \mathbf{u} = 0, \tag{4}$$

where $\mathbf{r}_i$ is the ith row of $\mathbf{W}_S$ and $\mathbf{u}$ is an unity row vector. For a minimum variance, distortionless response the $K \times 1$ vector $\mathbf{W}_C$ and the $KJ \times 1$ vector $\tilde{\mathbf{W}}_C$ are given by

$$\mathbf{W}_C = \frac{1}{K}\mathbf{u}, \qquad \tilde{\mathbf{W}}_C = \begin{bmatrix} \mathbf{W}_C \\ \cdots\cdot \\ \mathbf{0} \end{bmatrix}. \tag{5}$$

The GSC then implements the unconstrained LMS algorithm

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu y(k)\mathbf{X}_S(k), \tag{6}$$

where

$$\mathbf{X}_S(k) = [\mathbf{W}_S\mathbf{X}_1(k) \quad \mathbf{W}_S\mathbf{X}_2(k) \quad \cdots \quad \mathbf{W}_S\mathbf{X}_J(k)] = \tilde{\mathbf{W}}_S\mathbf{X}(k). \tag{7}$$

$\tilde{\mathbf{W}}_S$ is a $MJ \times KJ$ block diagonal matrix composed of J $\mathbf{W}_S$ matrices of dimension $M \times K$, and the error signal is the array output (since $d(k) = 0$ in Equation 1). Extensive details are provided in [6].

The LMS algorithm in Equation 6 is an iterative solution to the classic Weiner-Hopf equation for the optimal (in the mean-square sense) array weights shown in Figures 1 and 2. If $\mathbf{R}_{XX}$ represents the correlation matrix of the data vector $\mathbf{X}$ and $\mathbf{R}_{Xd}$ represents the cross-correlation vector between $\mathbf{X}$ and the desired signal d, then the Weiner-Hopf equation can be written

$$\mathbf{W}_{opt} = \mathbf{R}_{XX}^{-1}\mathbf{R}_{Xd}. \tag{8}$$

The optimal solution for the GSC weight vector is given by [4]

$$\mathbf{W}_{opt} = \left(\tilde{\mathbf{W}}_S\mathbf{R}_{XX}\tilde{\mathbf{W}}_S^T\right)^{-1}\tilde{\mathbf{W}}_S\mathbf{R}_{XX}\tilde{\mathbf{W}}_C, \tag{9}$$

which may also be written in the form of the Wiener-Hopf equation, Equation 8:

$$\mathbf{W}_{opt} = \mathbf{R}_{X_S}^{-1}\mathbf{R}_{Xd}, \tag{10}$$

with the GSC correlation matrix

$$\mathbf{R}_{X_S} = \tilde{\mathbf{W}}_S\mathbf{R}_{XX}\tilde{\mathbf{W}}_S^T \tag{11}$$

and the quiescent response vector

$$\mathbf{R}_{Xd} = \tilde{\mathbf{W}}_S\mathbf{R}_{XX}\tilde{\mathbf{W}}_C. \tag{12}$$

The eigenstructure of the $MJ \times MJ$-dimensional spatio-temporal correlation matrix $\mathbf{R}_{X_S}$ determines the speed of convergence of the standard LMS algorithm used in the GSC.

3 The Orthogonal LMS Processors

The stochastic gradient algorithms are a family of iterative search techniques for descending towards the optimal weights at a performance surface minimum. Another method of solving the Wiener-Hopf equation is to estimate the correlation matrix (or its inverse) and the cross-correlation vector at time k, and directly plug these estimates into Equation 8. This method of least squares is termed SMI and said to be deterministic or based on the observation data itself, while the LMS algorithm is said to be based on the observation statistics [7]. The least squares solution may be determined recursively using the matrix inversion lemma and is then termed the RLS algorithm. As will be shown, a better characterization of the least squares algorithm is that it is a form of LMS which requires that Equation 8 be satisfied at each iteration, while the conventional LMS requires this condition only in the limit.

Another equivalent description of RLS is that it incorporates a transformation of the input data such that the weight vector update points in the direction of the minimum mean-square error rather than in the direction of the negative gradient. For high eigenvalue disparity, the LMS algorithm (assuming that the initial weight vector is not on a principal axis) first walks in the direction of the negative gradient until it reaches a principal axis, and then proceeds towards the minimum, so that RLS can be considered more efficient.

3.1 Recursive Least Squares and the LMS-Newton Algorithm

The LMS-Newton algorithm [8] is equivalent to RLS, where LMS-Newton utilizes an a posteriori error estimate while RLS uses an a priori estimate. Both the least squares algorithms and LMS-Newton will be termed Orthogonal LMS (OLMS) algorithms, since they satisfy the normal equations iteratively.

The OLMS algorithms take the form

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \hat{\mathbf{R}}_{XX}^{-1}(k)\mathbf{X}(k)e(k), \tag{13}$$

where $e(k)$ is the error signal and $\mathbf{X}(k)e(k)$ is again the instantaneous estimate of the gradient. The issue at hand is the estimation of the correlation matrix inverse in Equation 13. Assuming ergodicity, if the correlation matrix estimate at time k utilizes an average of the input data vector outer products through time k, then the estimate is said to have infinite memory. It is also the best unbiased estimator. However, in a non-stationary environment, this would lead to poor tracking capabilities, since the memory of the last environment is embedded within the estimate. For this reason, an exponential weighting is often used to place more value on the most current terms and less value on the older memory.

The relationship of RLS to the LMS-Newton algorithm has been examined by Widrow [9], where both are referred to as first-order adaptive algorithms. Consider the derivation of the RLS algorithm utilizing the exponentially weighted averages

$$\hat{\mathbf{R}}_{XX}(k) = \sum_{l=1}^{k} \lambda^{k-l}\mathbf{X}(l)\mathbf{X}^T(l), \quad \hat{\mathbf{R}}_{Xd}(k) = \sum_{l=1}^{k} \lambda^{k-l}\mathbf{X}(l)d(l); \quad 0 \le \lambda \le 1, \tag{14}$$

where λ is the exponential weighting factor. The recursion of these estimates may be achieved by evaluating them at time $k-1$ and comparing the two forms to yield

$$\hat{\mathbf{R}}_{XX}(k) = \lambda\hat{\mathbf{R}}_{XX}(k-1) + \mathbf{X}(k)\mathbf{X}^T(k), \qquad \hat{\mathbf{R}}_{Xd}(k) = \lambda\hat{\mathbf{R}}_{Xd}(k-1) + \mathbf{X}(k)d(k). \quad (15)$$

The correlation matrix inverse may be recursively found via the matrix inversion lemma, where we define $\mathbf{P}(k) = \hat{\mathbf{R}}_{XX}^{-1}(k)$,

$$\mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) - \frac{\lambda^{-2}\mathbf{P}(k-1)\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{P}(k-1)}{1+\lambda^{-1}\mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k)}. \quad (16)$$

Introducing a vector which is often called the Kalman gain as

$$\mathbf{K}(k) = \frac{\lambda^{-1}\mathbf{P}(k-1)\mathbf{X}(k)}{1+\lambda^{-1}\mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k)}, \quad (17)$$

we find that the matrix inverse recursion becomes

$$\mathbf{P}(k) = \lambda^{-1}\mathbf{P}(k-1) + \lambda^{-1}\mathbf{K}(k)\mathbf{X}^T(k)\mathbf{P}(k-1). \quad (18)$$

Through the rearrangement of Equation 17 and comparison to Equation 18, it can be seen that

$$\mathbf{K}(k) = \mathbf{P}(k)\mathbf{X}(k) = \hat{\mathbf{R}}_{XX}^{-1}(k)\mathbf{X}(k). \quad (19)$$

The least squares weight vector update is then determined by the Wiener-Hopf equation

$$\begin{aligned} \mathbf{W}(k) &= \mathbf{P}(k)\hat{\mathbf{R}}_{Xd}(k) \\ &= \lambda\mathbf{P}(k)\hat{\mathbf{R}}_{Xd}(k-1) + \mathbf{P}(k)\mathbf{X}(k)d(k) \\ &= \mathbf{P}(k-1)\hat{\mathbf{R}}_{Xd}(k-1) - \mathbf{K}(k)\mathbf{X}^T(k)\mathbf{P}(k-1)\hat{\mathbf{R}}_{Xd}(k-1) + \mathbf{P}(k)\mathbf{X}(k)d(k) \\ &= \mathbf{W}(k-1) - \mathbf{K}(k)\mathbf{X}^T(k)\mathbf{W}(k-1) + \mathbf{P}(k)\mathbf{X}(k)d(k) \\ &= \mathbf{W}(k-1) + \mathbf{K}(k)\alpha(k), \end{aligned} \quad (20)$$

where the a priori error $\alpha(k)$ is given by

$$\alpha(k) = d(k) - \mathbf{X}^T(k)\mathbf{W}(k-1), \quad (21)$$

and via the substitution of Equation 19 into the last term of 20, we find

$$\mathbf{W}(k) = \mathbf{W}(k-1) + \hat{\mathbf{R}}_{XX}^{-1}(k)\mathbf{X}(k)\alpha(k). \quad (22)$$

Whether the weight vector in Equations 20 and 22 is determined at time k or $k+1$ by $\hat{\mathbf{R}}_{XX}^{-1}(k)$ and $\hat{\mathbf{R}}_{Xd}(k)$ is the difference between RLS and LMS-Newton; RLS chooses $\mathbf{W}(k)$ and LMS-Newton uses $\mathbf{W}(k+1)$. Thus the LMS-Newton algorithm takes the form

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \hat{\mathbf{R}}_{XX}^{-1}(k)\mathbf{X}(k)\varepsilon(k), \tag{23}$$

where the standard LMS a posteriori error estimate

$$\varepsilon(k) = d(k) - \mathbf{X}^T(k)\mathbf{W}(k) \tag{24}$$

is used to form the gradient estimate. Note that both Equations 22 and 23 are in exactly the same form as Equation 13.

The primary advantage of the OLMS algorithms is that they provide, under the assumptions of stationarity and ergodicity, better convergence characteristics than those of the standard LMS algorithm. This is due to the transformation of the observation data via its pre-multiplication by $\hat{\mathbf{R}}_{XX}^{-1}$. Assuming stationarity and ergodicity, the convergence behavior of OLMS algorithms may be summarized [7] as:

1. The weight vector is convergent in the mean and converges in the norm almost linearly with time.
2. The algorithm converges in the mean square in about $2M$ iterations, where M is the number of taps in the TDL.
3. The mean-square a priori error converges to the minimum mean-square error, yielding theoretically no misadjustment.

It is emphasized that the least squares, or OLMS, algorithms are derived from recursive forms of SMI which directly solve the Wiener-Hopf equation. The form of the solution is exactly a steered LMS, where the update is transformed via a rotational operator to point in the direction of minimum mean-squared error as opposed to that of the negative gradient.

Through the comparison of Equation 1 with Equations 22 and 23, it is clear that the sole difference between the OLMS algorithms and LMS is the pre-multiplication of the weight update correction term (the gradient estimate) by the inverse of the correlation matrix. As previously mentioned, this is a rotational operator. The improvement in the speed of convergence realized by OLMS algorithms and their relative invariance to the eigenvalue spread of the correlation matrix may now be explained. The LMS weight vector follows a trajectory proportional to $(\mathbf{I} - 2\mu\mathbf{R}_{XX})^n$ [1]. The OLMS algorithms replace μ by $\mathbf{R}_{XX}^{-1}$, so the trajectory theoretically converges independent of the eigenstructure of $\mathbf{R}_{XX}$.

3.2 The Karhunen-Loève OLMS Processor

The optimal decomposition of the matrix $\mathbf{R}_{X_s}$ (Equation 11) is achieved via the two-dimensional Karhunen-Loève Transform (KLT). The KLT is optimal in that it fully decorrelates the data by compacting the energy into the fewest spectral coefficients. Note that in order for this to be true the KLT operator must operate on the entire stacked vector $\mathbf{X}_S$.

To see the utility of the KLT, consider a Karhunen-Loève decomposition of the spatial coordinates. This would be accomplished by the $M \times M$-dimensional KLT operating upon each set of data vectors prior to their time-delay. The block denoted $\mathbf{Q}$ in Figure 3, performs the spatial KLT. This decomposition would ensure that the spatial dimension was decorrelated; however, nothing could be assumed about the temporal coordinates. The resulting spatio-spectral correlation matrix would have diagonal matrices in the J blocks, each of dimension $M \times M$, that formed the block-diagonal of $\mathbf{R}_{Z_s} = \mathbf{Q}_{sklt}\mathbf{R}_{X_s}\mathbf{Q}_{sklt}^T$ where $\mathbf{Q}_{sklt}$ represents the spatial KLT.

The KLT could also be performed only in the temporal dimension. This is depicted in Figure 4. If the equivalent spatio-temporal correlation matrix $\tilde{\mathbf{R}}_{X_S}$ was formed from row-stacked vectors, then the result of the temporal KLT would be the diagonalization of the M $J \times J$-dimensional blocks on the block-diagonal of $\tilde{\mathbf{R}}_{Z_t} = \mathbf{Q}_{tklt}\tilde{\mathbf{R}}_{X_s}\mathbf{Q}_{tklt}^T$ where $\mathbf{Q}_{tklt}$ represents the temporal KLT.

Now consider the complete two-dimensional KLT operator acting upon the $MJ \times MJ$-dimensional spatio-temporal correlation matrix $\mathbf{R}_{X_S}$. The result would be the complete decorrelation (diagonalization) of the resulting matrix

$$\mathbf{R}_{ZZ} = \mathbf{Q}_{klt}\mathbf{R}_{X_S}\mathbf{Q}_{klt}^T, \tag{25}$$

which is represented in the general form of Figure 5.

Since the desired signal is removed from the lower path of the GSC by the signal blocking matrix, the error signal is replaced by the array output in the LMS algorithm as described

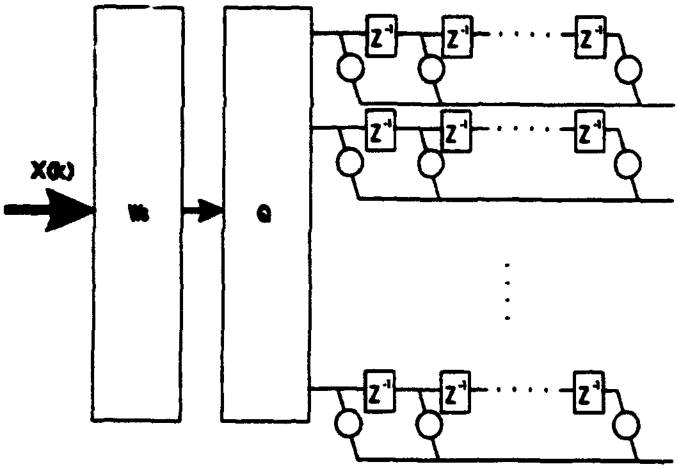


Figure 3: A spatial-KLT form GSC processor.

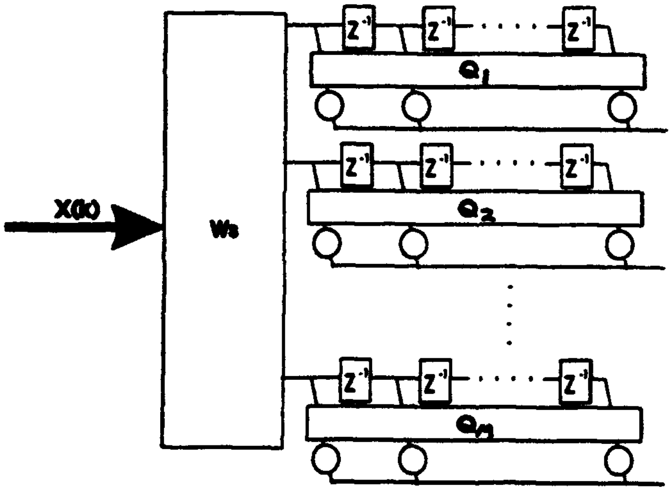

X(k)
Ws
Z⁻¹
Q1
Q2
QM

Figure 4: A temporal-KLT form GSC processor.

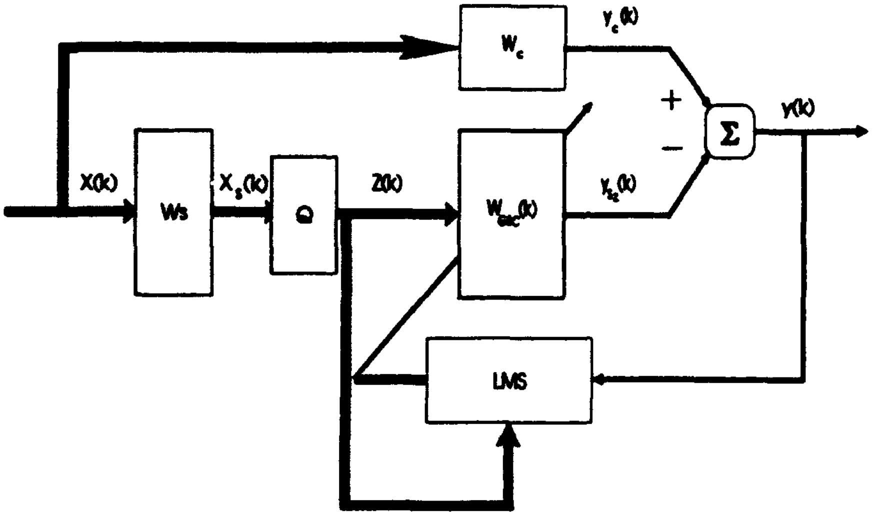


Figure 5: A general transform-domain form GSC processor.

in Section 2, and the corresponding GSC OLMS algorithm has the form

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{R}_{XX}^{-1}(k)\mathbf{X}_S(k)y(k). \tag{26}$$

The transformation

$$\mathbf{Z}(k) = \mathbf{Q}_{KLT}\mathbf{X}_S(k) \tag{27}$$

produces the diagonal correlation matrix in Equation 25, and the transformation domain OLMS algorithm takes the form

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mathbf{R}_{ZZ}^{-1}\mathbf{Z}(k)y(k). \tag{28}$$

The inversion of this correlation matrix corresponds to MJ scalar inversions. Hence, the computational complexity is greatly reduced for the matrix inverse operation, but the formation of the KLT is extremely complex. The OLMS algorithm is then on the computational order of LMS, plus the requirements of the observation data transform.

The update equation for the KLT OLMS in Equation 28 assumes exact knowledge of $\mathbf{R}_{ZZ}^{-1}$. This follows because the KLT transformation itself requires knowledge of $\mathbf{R}_{X_s}$, and $\mathbf{R}_{ZZ}$ is obtained from Equation 25.

If another orthogonal transformation Q is used, such as that implied by a Gram-Schmidt decomposition, then $\mathbf{R}_{ZZ}$ is not known a priori. Thus $\mathbf{R}_{ZZ}$ must be estimated. Since $\mathbf{R}_{ZZ}$

is diagonal (because $\mathbf{Q}$ decorrelates $\mathbf{X}_S$), then $\mathbf{R}_{ZZ}$ is estimated simply by estimating the average power in each element of $\mathbf{Z}$. $\hat{\mathbf{R}}_{ZZ}^{-1}$ is easily obtained with MJ scalar inversions.

3.3 $\hat{\mathbf{R}}_{ZZ}^{-1}$ Viewed as a Step-Size Matrix

It is noted that an alternative interpretation of Equation 28 is that of the standard LMS algorithm, where the transform has reduced the MJ-dimensional problem to MJ decoupled scalar LMS weight updates, and the inverse correlation matrix is a diagonal matrix of optimal step-sizes for each of the MJ modes. The KLT is unique in that it rotates the observation data to lie on the principal axis of the mean-square error surface. The eigenvalues then are a measure of the distance along the principal axes from the origin to the observation data. The pre-multiplication of the KLT transformed LMS weight update by $\mathbf{R}_{ZZ}^{-1}$ is then equivalent to choosing the step size for each decoupled mode such that one-step convergence is achieved.

In the case of an arbitrary orthogonal transformation $\mathbf{Q}$, $\hat{\mathbf{R}}_{ZZ}^{-1}$ in Equation 28 becomes a diagonal step size matrix $\mu(k)$ that we create using power estimates. The power estimates can be made using a single pole low-pass filter:

$$\hat{\sigma}_{Z_l}^2(k) = \beta\hat{\sigma}_{Z_l}^2(k-1) + (1-\beta)Z_l^2, \tag{29}$$

where β (such that $0 < \beta < 1$) controls the bandwidth of the filter and the resulting power averaging time. Then the lth diagonal element of $\mu(k)$ is

$$\mu_l(k) = \frac{\alpha}{\sigma_{Z_l}^2(k)}, \tag{30}$$

where α is a scalar. If the power estimates are exact, then the convergence rate of the algorithm is indicated by the eigenvalue spread, $\lambda_{max}/\lambda_{min}$, of the matrix $\mu(k)\mathbf{R}_{ZZ}$. When $\mathbf{Q}$ completely decorrelates $\mathbf{X}_S$, the eigenvalue spread is unity, which implies the fastest convergence.

The use of a time-varying step size was first considered by Griffiths [10], where a Gram-Schmidt (GS) decomposition was performed on the input data vector. This approach, however, does not provide much improvement in implementation because the overall complexity, due to the use of the GS decomposition, is on the order of that required by LS methods. We desire to obtain improved convergence speed without the need for such computational cost.

Both Compton [11] and Lee and Un [12] have examined the performance of the DFT frequency domain algorithm. Compton concluded that the TDL and DFT structures always perform identically in his analysis of single and multiple channel adaptive filters. His work, however, did not consider the Orthogonalized LMS algorithms, time-varying step sizes, or full two-dimensional transform. Lee and Un realized the possibility of achieving better convergence properties through the normalization of the step size, as have Narayan et al. [13]. However, both of the latter authors restricted their analysis to single channel filters.

3.4 Time-Varying Orthogonal Transforms

The direct implementation of the Gram-Schmidt algorithm serves as one method of realizing a completely orthogonal signal set to serve as the input to the adaptive processor [6]. The direct GS orthogonal structure using the LMS algorithm was first developed by Griffiths [10] and used a constant step size. This was later modified by Lee et al. [12] to include both a time-varying step size and an escalator realization, where the unit lower triangular transform factorization is used. Following the notation of [10], with the extensions of [12], the structure may be realized using a matrix **Q** which is composed of time-varying coefficients, and the MJ outputs satisfy

$$\mathbf{E}\left[z_m(k)z_n(k)\right] = 0, \quad m \neq n. \tag{31}$$

The matrix **Q** is lower triangular and composed of elements q_{ij} which may be represented as

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 \\ q_{2,1} & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot\ \cdot & & \cdot & \cdot \\ \cdot & \cdot\ \cdot & & \cdot & \cdot \\ \cdot & \cdot\ \cdot & & 1 & 0 \\ q_{(K-1)J,1} & \cdot\ \cdot & & q_{(K-1)J,(K-1)J-1} & 1 \end{bmatrix}. \tag{32}$$

The orthogonalization procedure generates the orthogonal output $\mathbf{Z}_m(k)$ via the recursive relationship

$$\begin{aligned} y_{m,1}(k) &= \hat{\mathbf{X}}_{S_m}(k) \\ y_{m,m}(k) &= z_m(k) \\ z_m(k) &= \hat{\mathbf{X}}_{S_m}(k) - \sum_{n=1}^{m-1} c_{m,n}(k)z_n(k); \quad 2 \leq m \leq (K-1)J, \end{aligned} \tag{33}$$

where the value of $c_{m,n}$ is chosen to minimize the local values of $\mathbf{E}\left[y^2_{m,n+1}(k)\right]$. In conjunction with the method of gradient descent, we may write

$$c_{m,n}(k+1) = c_{m,n}(k) + \mu \frac{\partial y^2_{i,j+1}(k)}{\partial c_{m,n}(k)}. \tag{34}$$

This result may be achieved through the use of the LMS algorithm to update the adaptive coefficients $c_{m,n}$

$$c_{m,n}(k+1) = c_{m,n}(k) + \mu_n(k) y_{m,n+1}(k) z_n(k), \tag{35}$$

where $\mu_n(k)$ is the time-varying step size formed in the manner described in Section 3.3. The matrix Q in Equation 32 is then given by $(\mathbf{I}+\mathbf{C})^{-1}$, where C is lower triangular with zeros on the diagonal and elements $c_{m,n}$. The orthogonality in the GS structure is complete after the convergence of the adaptive coefficients via the LMS algorithm in Equation 35. The GS decomposition requires $(MJ)(MJ-1)/2$ adaptive coefficients plus the MJ noise cancelling weights. Hence, the cost of the GS processor is exactly equal to that of SMI; and as Yuen [14] pointed out, so is the performance.

Griffiths [10] and Lee, et al. [12] also implemented the GS via a lattice filter structure. This structure decreased both the performance and the complexity. The lattice filter required convergence of PARCOR coefficients prior to providing an orthogonal output and had $MJ + 2M^2(J-1)$ adaptive coefficients requiring LMS updates.

4 Approximations to OLMS via Time-Invariant Orthogonal Transformations

The suboptimal transform of the observation data may not result in the complete decorrelation of the transformed data vector. Therefore $\mathbf{R}_{ZZ}$, and hence $\mathbf{R}^{-1}_{ZZ}$, may not be diagonal. The step size matrix $\mu(k)$ defined in Equation 30 may still be viewed as an estimator of $\mathbf{R}^{-1}_{ZZ}$, where, since $\mu(k)$ is diagonal, error is clearly incurred in the estimates of off diagonal elements. The decorrelating quality of the transform Q directly determines the quality of this estimator.

In this paper we indicate the decorrelating quality of Q with a quantity termed the efficiency η. η is defined as the eigenvalue spread of the matrix $\mathcal{M}\mathbf{R}_{\mathbf{z}\mathbf{z}}$, where

$$\mathcal{M} = \alpha \begin{bmatrix} \frac{1}{\hat{\sigma}^2_{\mathbf{z}_1}} & & & 0 \\ & \frac{1}{\hat{\sigma}^2_{\mathbf{z}_2}} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\hat{\sigma}^2_{\mathbf{z}_{K-1}}} \end{bmatrix}. \qquad (36)$$

Very early in the history of adaptive array research many investigators examined frequency domain LMS filters [15, 16, 17]. The frequency domain transformation is usually implemented with **Q** being an invertible matrix composed of the Discrete Fourier Transform (DFT) or the Discrete Cosine Transform (DCT) coefficients which operate in an identical manner upon each tapped delay line. The initial research in this area was limited to the analysis of the temporal transform LMS algorithm with a constant step size. Compton [11] then published a report which showed that this frequency domain structure performance was identical to that of the tapped-delay-line processor, again utilizing a fixed step size in the LMS algorithm. Subsequently, many other researchers began examining the use of transform domain adaptive filtering for narrowband single channel applications which considered the use of time-varying LMS step sizes [12, 13, 18, 19]. While the frequency domain has an intuitive appeal as a method of improving the performance of wideband adaptive arrays, it was pointed out in the preceding section that such a transformation only operates upon the temporal domain, and this transformation is sub-optimal for the two-dimensional data of interest. Hence, in this section we will consider the two-dimensional implementation of these transforms, resulting in a spatio-temporal frequency subbanding, which in effect reduces the wideband array problem to discrete frequency bins.

4.1 The Discrete Fourier and Discrete Cosine Transforms

The extension of transforming the observation data coupled with a time-varying LMS algorithm for array processing was first introduced using the DFT in [20]. The use of the DCT, motivated by the results in image and speech compression, was then presented for array processing in [6, 21, 22], where in [22], the DCT was implemented with a filter bank structure. All of these transforms only considered the temporal operator. The two-dimensional operator was first introduced in [23], with the DCT again providing the transformation.

4.1.1 Non-Separable Two-Dimensional Operators

The non-separable two-dimensional DFT is implemented upon the MJ-dimensional signals present prior to the weighting network according to

$$\mathbf{Z}_n = \frac{1}{\sqrt{MJ}} \sum_{m=1}^{MJ} \mathbf{X}_{S_m} e^{-j\left(\frac{2\pi(m-1)(n-1)}{MJ}\right)}, \tag{37}$$

or, equivalently, for the stacked MJ-dimensional data vector $\mathbf{X}_S$ at time k,

$$\mathbf{Z}(k) = \mathbf{Q}_{DFT}\mathbf{X}_S(k), \tag{38}$$

where $\mathbf{Q}_{DFT}$ is the $MJ \times MJ$ matrix of exponential coefficients which realize Equation 37. This algorithm has the additional benefit of not requiring an inverse transform to obtain the time domain output [24].

The Discrete Cosine Transform has the computational advantage of using only real numbers to provide a transform of the input data. Further, recent articles in the literature [24, 13, 25] reported that the narrowband single channel DCT adaptive filter provided better results than the DFT and other orthogonal transform filters for a class of data used in speech related applications. The non-separable two-dimensional DCT orthogonal transform is given by

$$\mathbf{Z}_n = \begin{cases} \frac{\sqrt{2}}{MJ} \sum_{m=1}^{MJ} \mathbf{X}_{S_m} & n = 1 \\ \frac{2}{MJ} \sum_{m=1}^{MJ} \mathbf{X}_{S_m} \cos\left(\frac{\pi(2m+1)(n+1)}{2MJ}\right) & n \neq 1, \end{cases} \tag{39}$$

which is represented at time k by the MJ-dimensional vector

$$\mathbf{Z}(k) = \mathbf{Q}_{DCT}\hat{\mathbf{X}}_S(k), \tag{40}$$

and $\mathbf{Q}_{DCT}$ is the $MJ \times MJ$ matrix of real coefficients which realize Equation 39.

4.1.2 Separable Two-Dimensional Operators

The tapped delay line of the GSC at time k has M spatial coordinates and J temporal coordinates. Hence the data may be represented by the $M \times J$ matrix χ. The two-dimensional transformation of this matrix may be accomplished by operating first upon the temporal dimension and then spatially upon the resulting matrix. Thus, defining the $M \times M$ spatial and $J \times J$ temporal DFT and DCT matrices, $\mathbf{Q}_{s_{DFT}}$, $\mathbf{Q}_{s_{DCT}}$, $\mathbf{Q}_{t_{DFT}}$, $\mathbf{Q}_{t_{DCT}}$, the transform domain matrices $\tilde{\mathbf{Z}}_{DFT}$ and $\tilde{\mathbf{Z}}_{DCT}$ are given by

$$\tilde{\mathbf{Z}}_{DFT} = \mathbf{Q}_{t_{DFT}} \chi^T \mathbf{Q}^T_{s_{DFT}}, \qquad \tilde{\mathbf{Z}}_{DCT} = \mathbf{Q}_{t_{DCT}} \chi^T \mathbf{Q}^T_{s_{DCT}}. \tag{41}$$

Column-stacking the matrices $\tilde{\mathbf{Z}}_{DFT}$ and $\tilde{\mathbf{Z}}_{DCT}$ yield the vectors $\mathbf{Z}_{DFT}$ and $\mathbf{Z}_{DCT}$, respectively. This operation may be simplified through the introduction of the Kronecker Products

$$\mathbf{Q}^s_{DFT} = \mathbf{Q}_{t_{DFT}} \otimes \mathbf{Q}_{s_{DFT}}, \tag{42}$$

and

$$\mathbf{Q}^s_{DCT} = \mathbf{Q}_{t_{DCT}} \otimes \mathbf{Q}_{s_{DCT}}, \tag{43}$$

where the superscript s denotes separability and the $\otimes$ operator denotes the standard Kronecker product. Thus

$$\mathbf{Z}^s_{DFT}(k) = \mathbf{Q}^s_{DFT} \mathbf{X}_S(k) \text{ and } \mathbf{Z}^s_{DCT}(k) = \mathbf{Q}^s_{DCT} \mathbf{X}_S(k). \tag{44}$$

4.2 Dependence on Signal Environment

For any time-invariant transform, the partitioning of the spatio-temporal frequency domain is fixed in the selection of the transformation. Since ideal filters are not realizable, there will be sidelobes in the passbands of adjacent filters. This spatio-spectral frequency "bleeding" prevents the diagonalization of the transform domain correlation matrix if sufficient jammer power is present in the sidelobes of adjacent filters. That is, the data in the spatio-spectral bins will be correlated. Thus, the efficiency η of the transform is a function of the jammer signal characteristics. This undesireable trait is not correctable after the selection of the processor transformation.

5 Simulation Results

We desire to find transforms which exhibit good performance improvements for a wide range of signal environments while maintaining computational simplicity. To that end we consider the DFT and DCT, both in separable and non-separable form, in the next section. We assume a single desired signal and allow a single jammer to move through a range of frequencies and directions of arrival in order to see how well each transform works for this

Table 1: Simulated Signal Environment

Signal	DOA	Center Frequency	Bandwidth	Power
desired	0°	$0.25f_n$	10%	-10 dB
jammer	variable $(-90°$ to $90°)$	variable (0 to $0.5f_n$)	10%	1 dB

class of environments. At each jammer frequency and angle of arrival the environment is assumed stationary and the convergence of the adaptive weights is examined by plotting the efficiency, or the eigenvalue spread, of the transformed matrix $\mathcal{M}\mathbf{R}_{ZZ}$.

The analysis of a GSC form array composed of $K = 4$ sensors and $J = 4$ taps per filter is considered. The tapped delay line has a spacing of Δ. Therefore, we will use notation of a normalized frequency f_n, where $f_n = 1$ indicates a frequency of $1/\Delta$ Hertz. The signal geometry was designed such that one desired signal and one jammer impinge upon the array, where the propagating signals' center frequencies, bandwidths, and powers are presented in Table 1.

The $MJ \times MJ$-dimensional DCT and DFT matrices were determined for the non-separable and separable cases. The non-separable transform matrices were calculated directly using the standard MJ-dimensional transforms on the stacked data vector $\mathbf{X}_s(k)$. The separable transform matrices were calculated by first finding the $M \times M$ spatial and $J \times J$ temporal transform matrices $\mathbf{Q}_s$ and $\mathbf{Q}_t$, respectively. Then, the separable $\mathbf{Q}^s_{DCT}$ and $\mathbf{Q}^s_{DFT}$ are formed by using the Kronecker product of the temporal and spatial transform matrices, given by $\mathbf{Q}^s_{DCT} = \mathbf{Q}_{t_{DCT}} \otimes \mathbf{Q}_{s_{DCT}}$ and $\mathbf{Q}^s_{DFT} = \mathbf{Q}_{t_{DFT}} \otimes \mathbf{Q}_{s_{DFT}}$, where "$\otimes$" denotes the Kronecker product.

Figure 7 displays the eigenvalue spread of the correlation matrix $\mathbf{R}_{X_s}$. The eigenvalue spread is minimal at 0 degree arrival angle due to the fact that the desired signal and the jammer are both impinging the array at 0 degrees. This can be explained by considering the mean square error (MSE) of the GSC shown in Figure 6. The MSE is given by

$$E[y^2] = (\tilde{\mathbf{W}}_C^T - \mathbf{W}_{opt}^T \tilde{\mathbf{W}}_S)\mathbf{R}_{xx}(\tilde{\mathbf{W}}_C^T - \mathbf{W}_{opt}^T \tilde{\mathbf{W}}_S)^T, \tag{45}$$

where y is the array output. The MSE is the same for all transforms. We observe in Figure

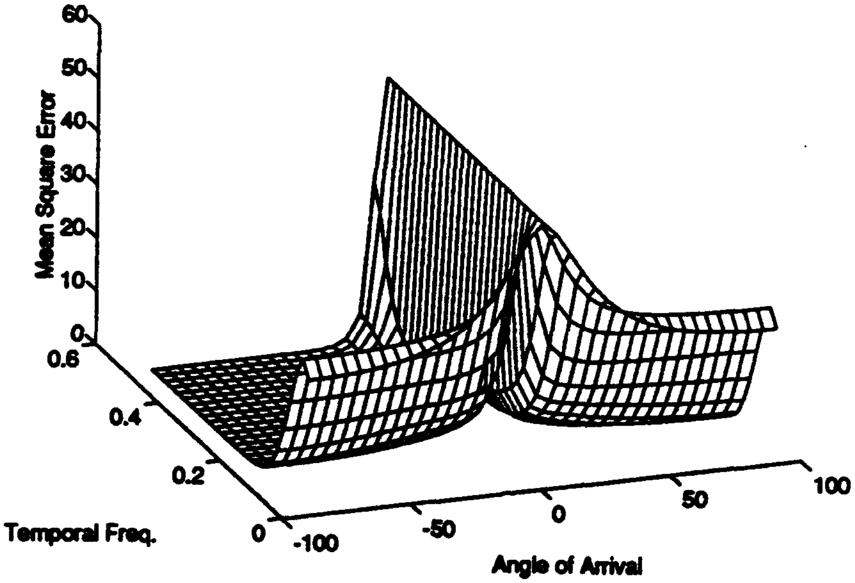


Figure 6: MSE (Applicable to all the Transforms).

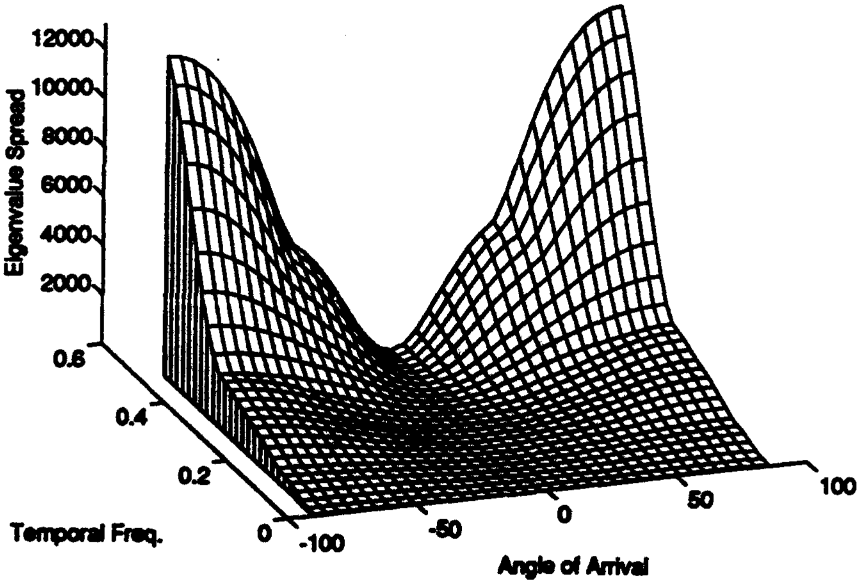


Figure 7: Eigenvalue Spread, Identity transform.

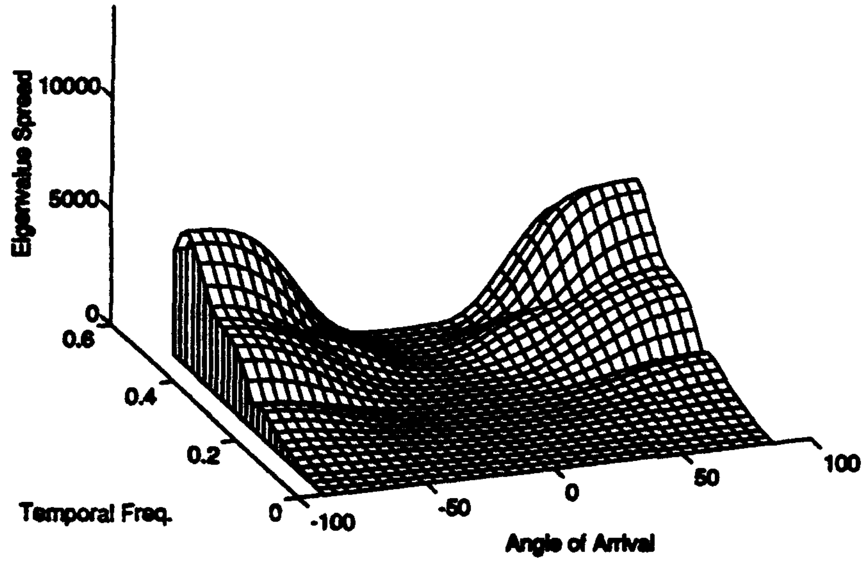


Figure 8: Eigenvalue Spread, Separable DFT.

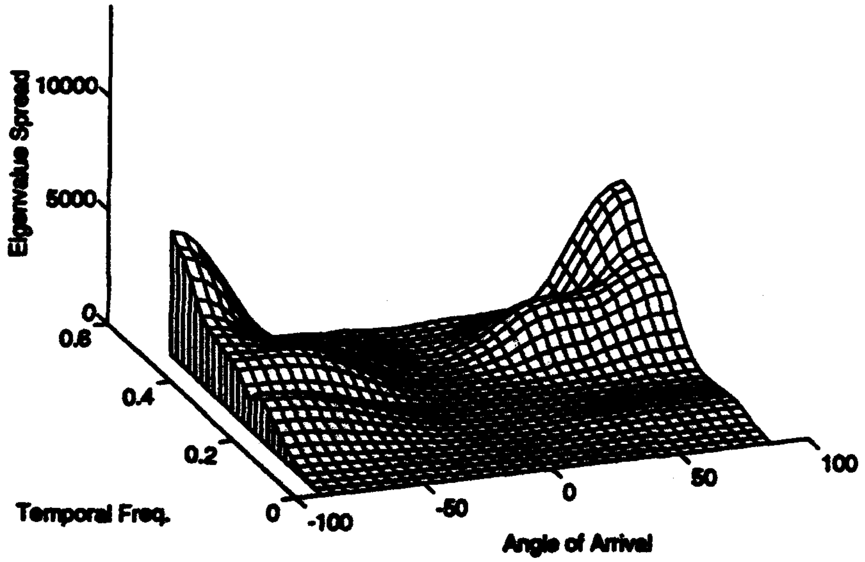


Figure 9: Eigenvalue Spread, Non-separable DFT.

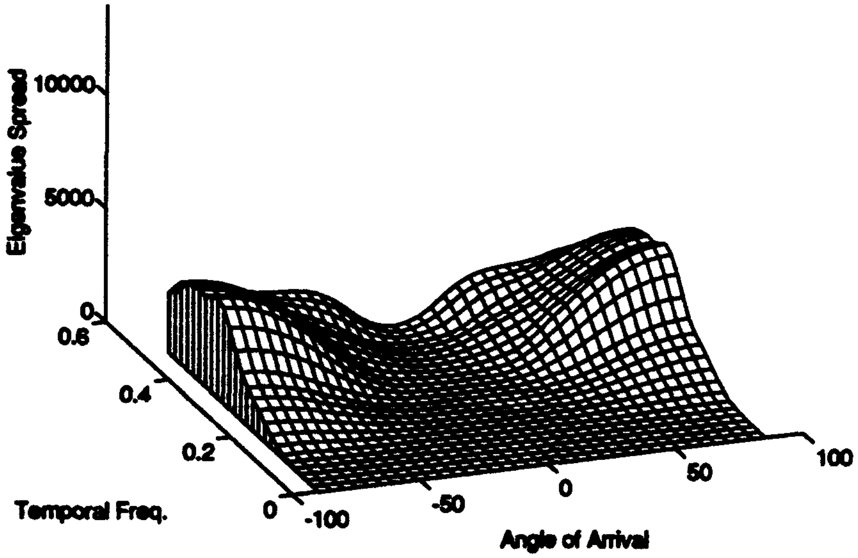


Figure 10: Eigenvalue Spread, Separable DCT.

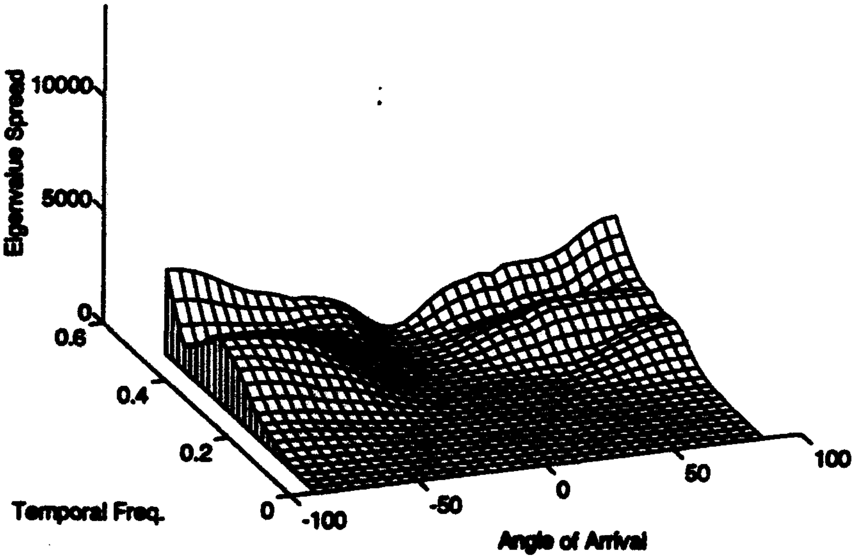


Figure 11: Eigenvalue Spread, Non-separable DCT.

6 that the MSE is a maximum at close to 0 degrees angle of arrival of the jammer. This indicates that the array fails to distinguish between the jammer and the desired signal in this region. Thus, both the jammer and desired signal are blocked and only noise appears in the correlation matrix $\mathbf{R}_{X_s}$. The resulting eigenvalue spread is minimal, but also meaningless.

The maximum eigenvalue spread in Figure 7 occurs in the regions where the temporal frequencies close to $0.5f_n$ and angles of arrival near ±90 degrees. Figures 8 - 11 display the performance measures of the transformed correlation matrices. It is observed that the transformations significantly reduce the maximum eigenvalue spread. While the non-separable transforms seem to give the best performance of eigenvalue spread reduction over the separable transforms, both give significant improvement over the identity operator. The DCT also outperforms the DFT for both separable and non-separable transforms.

6 Conclusions

The transformation of the observation data prior to adaptive processing provides the possibility of increasing the speed of convergence of the LMS algorithm with a time-varying step size. If the transformation is capable of completely decorrelating the observation data, then the performance of OLMS algorithms is obtained. The Gram-Schmidt LMS decomposition provides one link between OLMS and LMS, since GS LMS yields cost and performance identical to that of OLMS. The use of a time-invariant transformation reduces the cost, and the performance of the adaptive processor becomes dependent upon the efficiency of the transformation. This efficiency is environment dependent, so the choice of the transformation is important. Because the transformation conserves power, it does not by itself alter the eigenstructure of the correlation matrix. It is only by combining the time-varying step-sizes, obtained using estimates of the signal power in spatio-spectral bins, with these transformations that a performance improvement can be realized.

The efficiency of the transformation applied to the observation data determines the quality of the diagonal inverse correlation matrix estimate in the OLMS algorithms. If the transform completely decorrelates the observation data, then the transform is efficient and the quality of $\hat{\mathbf{R}}_{XX}^{-1}$ is excellent. If the transform does not completely decorrelate the observation data, then the quality of $\hat{\mathbf{R}}_{XX}^{-1}$ depends on the degree to which the observation data is decorrelated.

The time-invariant orthogonal transforms play an important role in this regard. In fact, the standards for compression in imaging and video are based on the DCT time-invariant transformation due to its good decorrelation properties over a wide range of processes and the fact that fast transforms exist for its computation [26].

Within the context of array processing, however, the efficiency of the time-invariant transforms are a function of the jammer; one of the variables over which the array designers have no control. It has been shown that the DFT and DCT, in general, are capable of reducing the eigenvalue spread of the resulting correlation matrix when coupled with a time-varying stepsize. The non-separable DCT provided the best performance of the transforms considered.

7 Acknowledgments

The authors take pleasure in recognizing the establishment of Krystal's for supplying sustenance, if not nutrition, during the final hours of our sanity. In addition, we are pleased to recognize Jim Dorsey's accomplishments from which we received motivation and inspiration. Finally, we desire to thank John Dorsey for his patience and selflessness in accepting his wife's long term absence.

References

[1] B. Widrow and M.E. Hoff, "Adaptive switching circuits," *IRE WESTCON Record*, Pt. 4, pp. 96–104, 1960.

[2] C.A. Baird and J.T. Rickard, "Recursive estimation in array processing," *Fifth Asilomar Conf on Circuits and Systems*, Pacific Grove, CA, pp. 509–513, Nov. 1971.

[3] I.S. Reed, J.D. Mallett and L.E. Brennan, "Sample-matrix inversion technique," *Proc. 1974 Adaptive Antenna Workshop*, NRL Report 7803, Vol. I, Washington, D.C., pp. 219–222, March 1974.

[4] L.J. Griffiths and C. W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Trans. Antennas and Propagat.*, Vol. AP-30, No. 1, pp. 27-34, January 1982.

[5] O.L. Frost, "An Algorithm for Linearly Constrained Adaptive Array Processing," *Proc. IEEE*, Vol. 60, No. 8, pp. 926-935, August 1972.

[6] J.S. Goldstein, *The Dynamic Behavior of Constrained Adaptive Array Sensor Processors*, USAF Rome Laboratory Technical Report RL-TR-92-327, December 1992. ADA262402

[7] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall: Englewood Cliffs, NJ, 1991.

[8] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall: Englewood Cliffs, NJ, 1985.

[9] B. Widrow, W. Newman, R. Gooch, K. Duvall, and D. Shur, *Research on Algorithms for Adaptive Array Antennas*, USAF Rome Laboratory Technical Report RADC-TR-81-206, August 1981. **ADA106684**

[10] L. J. Griffiths, "Adaptive Structures for Multiple-Input Noise Cancelling Applications," in *Proc. ICASSP*, pp. 925-928, April 1979.

[11] R.T. Compton, *On the Equivalence Between Tapped Delay-Line and FFT Processing in Adaptive Arrays*, Ohio State University ElectroScience Laboratory, Final Technical Report 717253-5, June 1986.

[12] J.C. Lee and C. K. Un, "Performance of Transform-Domain LMS Adaptive Filters," *IEEE Trans. Acoust., Speech, Sig. Proc.*, Vol. ASSP-34, No. 3, pp. 499-510, June 1986.

[13] S.S. Narayan, A.M. Peterson, and M.J. Narashima, "Transform Domain LMS Algorithm," *IEEE Trans. Acoust., Speech, Sig. Proc.*, Vol. ASSP-31, No. 3, pp. 609-615, June 1983.

[14] S.M. Yuen, "Comments on convergence properties of Gram-Schmidt and SMI adaptive algorithms," *IEEE Trans. Aerospace and Electronic Systems*, Vol. 27, No. 6, pp. 897–899, November 1991.

[15] B. Widrow, J. McCool, and M. Ball, "The Complex LMS Algorithm," *Proc. IEEE*, Vol. 63, No. 4, pp. 719-720, April 1975.

[16] M. Dentino, J. McCool, and B. Widrow, "Adaptive Filtering in the Frequency Domain," *Proc. IEEE*, Vol. 66, No. 12, pp. 1658-1659, Dec. 1978.

[17] N.J. Bershad and P.L. Feintuch, "Analysis of the Frequency Domain Adaptive Filter," *Proc. IEEE*, Vol. 67, No. 12, pp. 1658-1659, Dec. 1979.

[18] G.A. Clark, M.A. Soderstrand, and T.G. Johnson, "Transform Domain Adaptive Filtering Using a Recursive DFT," in *Proc. ISCAS*, pp. 1113-1116, 1985.

[19] N.J. Bershad and P.L. Feintuch, "A Weighted Normalized Frequency Domain LMS Adaptive Algorithm," *IEEE Trans. ASSP*, Vol. ASSP-34, No. 3, pp. 452-461, June 1986.

[20] J.S. Goldstein, "A Linearly Constrained Wideband Adaptive Array Antenna with Orthogonal Filter Structure," in *Proc. IEEE Intl. Symp. Antennas Propagat.*,Vol. I, pp. 612-615, July 1992.

[21] J.S. Goldstein, "A Wideband Multichannel Array Sensor with a Frequency Domain Adaptive Filter Structure," in *Proc. ICASSP*, Vol. IV, pp. 520-523, April 1993.

[22] J.S. Goldstein and M.A. Ingram, "A Wideband Sensor Array with an Adaptive Filter Bank Structure," in *Proc. IEEE Conf. Dual-Use Technol. and Applic.*, pp. 214-218, May 1993.

[23] J.S. Goldstein and M.A. Ingram, "Reduced complexity adaptive structures for jam-resistant satellite communications," to appear at MILCOM 93.

[24] N. Ahmed and K. R. Rao, *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag: Berlin, GDR, 1975.

[25] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. Computers*, p. 90-93, January, 1974.

[26] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall: Englewood Cliffs, NJ, 1989.

A PICOSECOND PULSE MEASUREMENT SYSTEM FOR NONDESTRUCTIVE EVALUATION (NDE) USING STEP-FREQUENCY RADAR

William H. Weedon and Weng Cho Chew
Department of Electrical and Computer Engineering
University of Illinois, Urbana, IL 61801

Abstract

A prototype step frequency radar measurement system is used to synthetically generate picosecond pulse measurement data. This system is useful for the nondestructive evaluation of many civil structures. Quantitative images of the scattering objects may be generated from the time-domain scattering data by employing a nonlinear inverse scattering algorithm. This algorithm can potentially generate a very high quality reconstruction of the scatterer with high resolution. However, proper calibration in the measurement system is critical to the success of this technique. We discuss three separate types of calibration used in the system. Finally, some reconstructions of scattering objects from measurement data collected with our system are shown.

1. Introduction

A prototype step frequency radar (SFR) system has been developed to synthetically generate picosecond pulse data similar what would be obtained from an impulse radar. The intended application is for nondestructive evaluation (NDE) of civil structures such as bridges, roads, tunnels, dams, buildings and aircraft runways. A nonlinear inverse scattering algorithm is used to process the microwave data and reconstruct a quantitative image of the scatterer from the measurement data.

Our nonlinear inverse scattering algorithm generates a computer model of the scattering objects in an iterative fashion and must solve a forward scattering problem at each iteration based on the computer model [1]. This is a very accurate and robust method of reconstructing the object and can generate a very high resolution image. However, high quality measurement data is required in order for the full capabilities of our reconstruction algorithm to be realized [2]. Therefore, calibration is a critical issue in the design of this system.

There are many calibration steps that must be performed in order to generate high quality data. We roughly categorize the calibration into three separate and distinct stages. The first is the calibration of the measurement itself in order to account for reflections, losses and phase changes in various components of the measurement system including transmission lines, connectors, directional couplers, switches, etc. The second calibration stage is the determination of the transfer function, or impulse response of the antennas used. Finally, we must perform a calibration between our computed scattering solution and the measurement data. We discuss these three calibration stages in more detail below.

The inverse scattering algorithm used is discussed only briefly here. For a detailed description of the inverse scattering algorithm and a comparison with other methods, see [1, 3, 4, 5]. Instead, attention is focused on the measurement system design and calibration issues. Both bistatic and monostatic measurement configurations may be used and we compare the two approaches. Finally, we show some object reconstructions from data collected with our prototype measurement system.

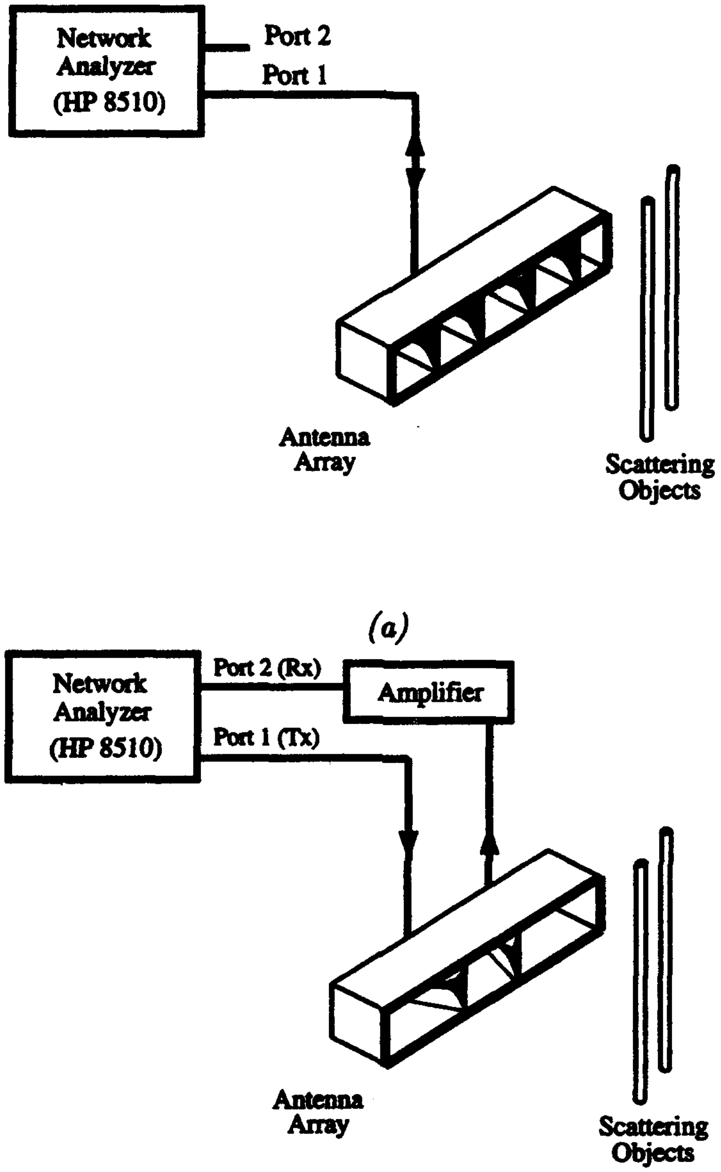
2. Overview of Measurement System and Data Processing

An HP-8510B vector network analyzer is the heart of our measurement system. The network analyzer servers as both transmitter and receiver, separates incoming from outgoing waves, and also performs various signal processing functions such as signal averaging. Data is collected by stepping the continuous wave source and receiver electronics from 2.0 GHz to 12.0 GHz in steps of 44.4 MHz for a total of 225 measured frequencies. Picosecond pulse measurement data is then generated from the SFR data by performing an inverse Fourier transform on the measured data.

A sketch of our prototype SFR measurement system for collecting monostatic measurement data is shown in Figure 1(a). This system consists of the HP-8510B network analyzer and an antenna array consisting of several broadband antennas. The antenna array is connected to port-1 of the network analyzer via a coaxial transmission line and port-2 is not used. In practice, we use a single antenna here and manually move it between various locations in the array.

The bistatic configuration is shown in Figure 1(b). For this case, two antennas are used and are both moved between various locations in the array. The transmitting antenna is connected to port-1 of the network analyzer while the receiving antenna is connected to port-2. An amplifier is inserted into port-2 between the receiving antenna and the network analyzer in order to improve the signal-to-noise ratio.

A block diagram of the various steps in the data processing used to reconstruct the scattering object from the measurement data is shown in Figure 2. The switch on the left allows the choice of either measured scattering data or computer generated data that is used to test the reconstruction algorithm. The algorithm begins with specification of the initial parameters, which are usually set to zero. Using the current computer model, forward scattering data is generated and subtracted from the measured data. This difference is then used to compute a measure of the residual field error. If the difference


Network
Analyzer
(HP 8510)
Port 2
Port 1
Antenna
Array
Scattering
Objects
Network
Analyzer
(HP 8510)
Port 2 (Rx)
Amplifier
Port 1 (Tx)
Antenna
Array
Scattering
Objects

(a)

(b)

Figure 1: Sketch of prototype SFR system for collecting (a) monostatic and (b) bistatic measurement data.

is below a specified tolerance, the current model parameters are displayed on a graphics workstation. If the field error is not below a specified tolerance, the field error is sent to a conjugate gradient optimization procedure which returns an update to the model parameters. The process is repeated until a

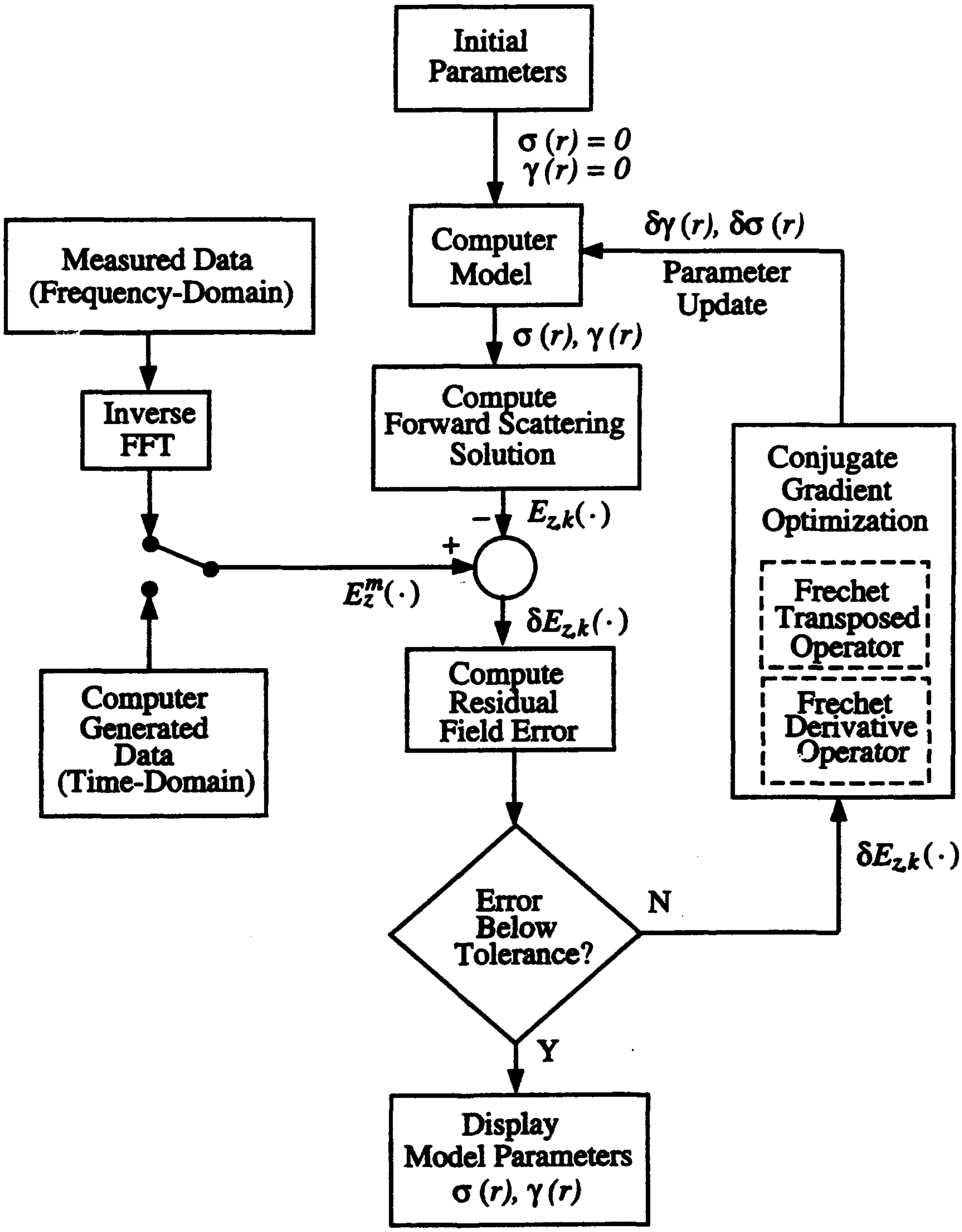

Initial Parameters
σ (r) = 0
γ (r) = 0
Computer Model
δγ (r), δσ (r)
Parameter Update
Measured Data (Frequency-Domain)
Inverse FFT
σ (r), γ (r)
Compute Forward Scattering Solution
Conjugate Gradient Optimization
Frechet Transposed Operator
Frechet Derivative Operator
−
E_{z,k}(·)
+
E_z^m(·)
δE_{z,k}(·)
Computer Generated Data (Time-Domain)
Compute Residual Field Error
δE_{z,k}(·)
Error Below Tolerance?
N
Y
Display Model Parameters σ (r), γ (r)

Figure 2: Block diagram of processing of measured and computer simulated scattering data.

convergent solution is attained.

A layout of the transmitters, receivers and object grid corresponding to

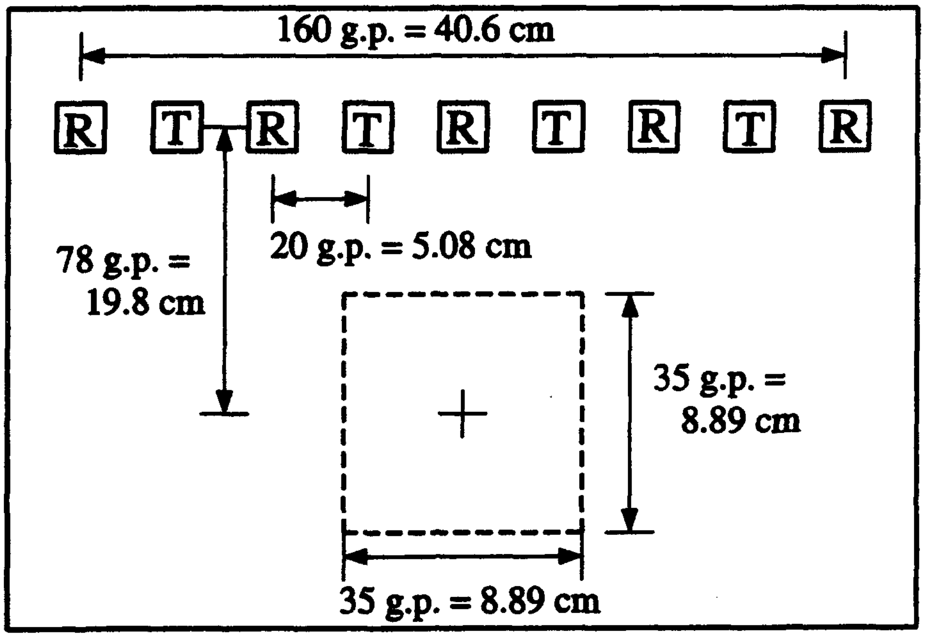


Figure 3: Arrangement of transmitters, receivers and object grid in bistatic measurement.

the bistatic measurement system in Figure 1(b) is shown in Figure 3. The configuration for the monostatic case is the same except that the transmitter (T) and receiver (R) locations in Figure 3 are replaced with transceivers. We assume a 2-D model for this prototype system. That is, we assume that the transmitters are infinite line sources and that the scattering objects extend to infinity, normal to the page.

3. Measurement System Calibration

For monostatic measurements, the measured field data is obtained from an S_{11} network analyzer measurement. We first take a field measurement with no scatterers in place and then take a second measurement with the scatterers. The scattered field is the difference of these two measurements.

There are several ways to calibrate an S_{11} measurement. The most common is perhaps the 3-term calibration model using a short, open and load. This nonlinear model is a very accurate method for accounting for the directivity of the directional couplers, frequency tracking that includes factors such as the gain of the transmission lines, and source match. However, this nonlinear calibration model is very sensitive to changes in the measurement system. We found this is not a very good calibration method for our purposes because we were forced to use flexible transmission line to connect the network analyzer to the antenna array. The process of switching antenna elements requires connecting, disconnecting and moving the transmission line. These small changes in the transmission line were enough to cause large errors in the measurement data when a nonlinear calibration model was used.

Instead of using a 3-term nonlinear calibration model, we instead chose to use a 1-term linear model that accounts for frequency tracking only. For monostatic measurements, the calibration consists of simply connecting a short to port-1, measuring the transfer function with the short in place, and dividing the real measurements by the negative of the transfer function with the short in place.

For bistatic measurements, a very accurate calibration method is to use a 6-term model consisting of short-open-load measurements on each port and thru measurements in each direction. However, this 2-port nonlinear calibration suffers from the same problems that we encountered with the monostatic measurements. Instead, we use a tracking calibration that is obtained by connecting ports 1 and 2 together with a thru connector and determining the transfer function. The real measurements are then connected by dividing by the thru measurement. This tracking calibration would, of course, account for the gain of the amplifier on port-2.

We suffer a loss in dynamic range by choosing a linear calibration procedure over a nonlinear one because we do not account for many systematic errors such as losses in the directional couplers and multiple reflections between the source and antenna input. However, our method is much more robust with respect to changes in the system and the calibration procedure takes much less time to perform. Our approach is therefore to use a very simple calibration procedure, but use high quality system components that have a minimal amount of reflections and losses.

4. Determination of the Antenna Impulse Response

Another important calibration procedure is in determining the impulse response of our antennas. Since we are planning on adding more transmitter and receiver antenna elements to our antenna array in the future, we wish to determine the impulse response of each antenna individually, rather than the cascade of two antennas. For narrowband antenna gain measurements, one would measure the response of the antenna under test (AUT) and compare the measurement to that of a standard gain pyramidal horn antenna. Measuring the impulse response of a broadband antenna is much more difficult because it requires determining the antenna gain over a wide bandwidth. The problem is that it is difficult to find a single standard gain antenna to cover the entire bandwidth. For example, the gain measurement of our broadband antenna from 2 GHz to 12 GHz in our laboratory would require at least 5 separate standard gain antennas and we would still have gaps in the spectrum.

To overcome the difficulties of finding standard gain horns to cover the bandwidth of our antenna, we devised a different procedure. We instead used a matched pair of 2–18 GHz conical horn antennas. The new procedure requires three separate measurements. The first involves connecting port-1 and port-2 together inside the antenna range with a long length (20 feet) of low loss coaxial semi-rigid cable. Next, a second measurement was made with the two matched conical horns in place. This allowed us to determine the transfer function of a single conical horn. Finally, a third measurement was made with a conical horn on one port and the AUT on the other port. From this measurement we were able to determine the transfer function of our AUT.

5. Calibration Between Computed Scattering Solution and Measurement

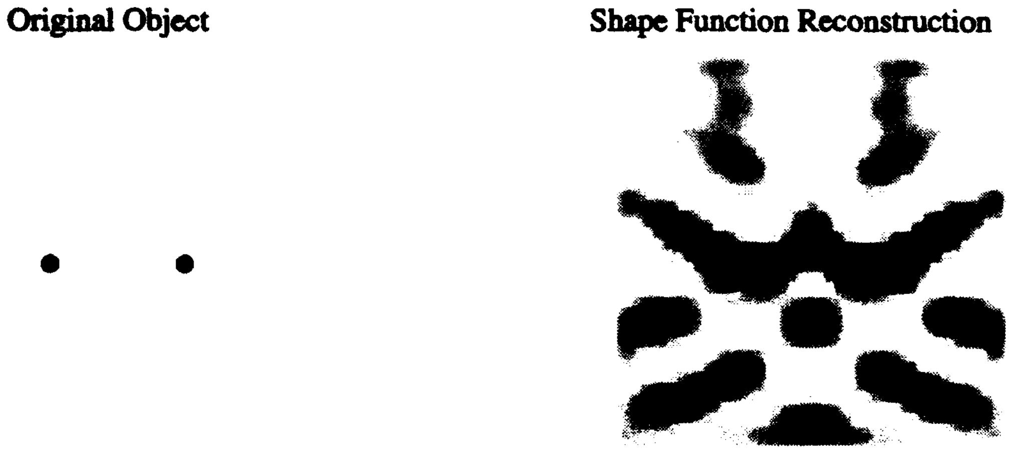
The third calibration procedure that we must perform is that between the computed scattering solution and the measurement data. This agreement must of course hinge on the proper modelling of the problem geometry and physics. Due to various modelling errors, there is usually an arbitrary gain constant (independent of frequency) and time offset common to all elements to be added in order make the computed solution agree with the measured data. Presently, we adjust these gain and offset constants by measuring the scattering from a known scatterer at a fixed position and tuning these parameters until the computed solution agrees with the true solution. However, we are searching for a better method to perform this calibration.

6. Reconstructions of Data Collected with Prototype System

To demonstrate the capabilities of our prototype NDE measurement system, we collected and processed the microwave scattering data from a bistatic measurement of two metallic cylinders. The cylinders both had a diameter of 1.0 cm and were separated by 3.2 cm. In Figure 4 we show the reconstruction for the case when the cylinders were aligned horizontally and in Figure 5 we show the case where the cylinders are aligned vertically. In both cases, the two cylinders are well resolved and the result is a high quality image of the two scatterers.

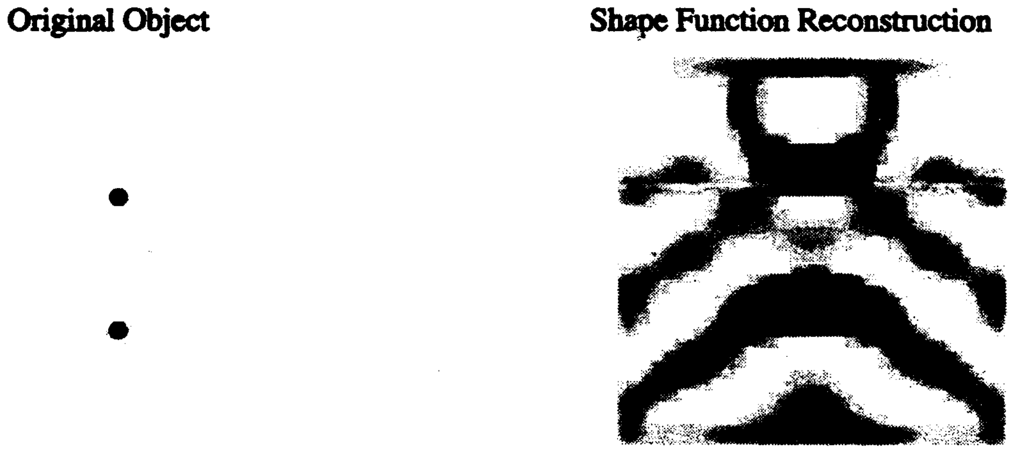
7. Conclusion

A prototype microwave measurement system has been developed using step frequency radar to synthetically generate picosecond pulse scattering


Original Object

Shape Function Reconstruction

Figure 4: Original object and shape function reconstruction of two metallic cylinders of diameter 1.0 cm aligned horizontally with separation 3.2 cm.


Original Object

Shape Function Reconstruction

Figure 5: Original object and shape function reconstruction of two metallic cylinders of diameter 1.0 cm aligned vertically with separation 3.2 cm.

data. The measured data was processed using a nonlinear inverse scattering algorithm and high resolution images of the scattering objects were obtained. Our prototype system and preliminary results allowed us to understand many calibration issues in the design of such a system.

Our results also demonstrate the capabilities of microwave technology and nonlinear inverse scattering as a means of performing NDE on certain

civil structures. We plan to further develop both our measurement system and inverse scattering algorithms in order to image larger objects with a quicker turnaround time.

Acknowledgement

This work is supported by the Office of Naval Research under grant N000-14-89-J1286, the Army Research Office under contract DAAL03-91-G-0339 and a grant from Schlumberger. Computer time is provided by the National Center for Supercomputer Applications at the University of Illinois, Urbana-Champaign.

The authors gratefully acknowledge the assistance of Professor Paul Mayes for providing the broadband antennas and associated expertise used in the experiment. We also thank Professor Jose Schutt-Aine for help with the calibration and use of the network analyzer. Finally, we thank our undergraduate student Chad Ruwe for providing help with various parts of the experiment.

References

[1] W. H. Weedon and W. C. Chew, "Time-domain inverse scattering using the local shape function (LSF) method," *Inverse Prob.*, vol. 8, pp. 1–14, 1993.

[2] W. H. Weedon, J. E. Mast, W. C. Chew, H. Lee, and J. P. Murtha, "Inversion of real transient radar data using the distorted-Born iterative algorithm," in *Proceedings of the IEEE Antennas and Propagation Society International Symposium*, (Chicago, IL), July 18–25, 1992.

[3] M. Moghaddam, W. C. Chew, and M. Oristaglio, "Comparison of the Born iterative method and Tarantola's method for an electromagnetic time-domain inverse problem," *Int. J. Imaging Systems Tech.*, vol. 3,

pp. 318–333, 1991.

[4] W. C. Chew and G. P. Otto, "Microwave imaging of multiple conducting cylinders using local shape functions," *IEEE Microwave Guided Wave Lett.*, vol. 2, pp. 284–286, July 1992.

[5] G. P. Otto and W. C. Chew, "Microwave inverse scattering–local shape function imaging for improved resolution of strong scatterers," *IEEE Trans. Microwave Theory Tech.*, 1993. Accepted for publication.

MONTE CARLO TECHNIQUES FOR PHASED ARRAY ANALYSIS

Dean A. Paschen and Steven C. Olson
Ball Communication Systems Division
Broomfield, CO 80038

ABSTRACT

Analysis of the effect of errors in phased array antennas is very well understood for special cases. Closed form expressions, curves, and/or tables have been presented in the literature to cover these special cases. However, simple expressions are not available for all combinations of these effects. With the increasing performance of computers, including desktop systems, Monte Carlo analysis techniques can now be used to quickly develop models for any combination of error conditions. The individual contributions of each effect can be isolated from the others and compared in statistical fashion as an aid to optimizing the algorithms used in the operation of the phased array. This type of analysis technique is also useful in showing the effects of long term degradation to the system due to component failures.

Specific examples from three different types of arrays are presented to show the benefit of Monte Carlo analysis techniques. A multiple beam phased array is analyzed to determine optimum quantization level and function for amplitude and phase, optimum steering algorithm for forming the multiple beams and minimizing interaction between beams, and composite performance with all effects combined. Analysis of random and systematic errors in components are shown for a radiometer antenna. Finally, the beam steering algorithm for a spherical switched beam antenna is optimized with the use of Monte Carlo methods, and the component tolerances are specified based on the same analysis.

1. INTRODUCTION

Error effects in phased array antennas can be divided into two types: systematic and random. Quantization errors are a type of systematic error which create larger sidelobes (quantization lobes) at certain angles relative to the main beam of the array. Random errors, typically caused by tolerances present in the fabrication of the array and its components, increase the energy in the sidelobe region and thereby reduce the main beam gain. This paper presents examples of Monte Carlo techniques for analyzing these types of errors.

The effect of quantization errors has previously been analyzed in closed form for single beam phased array antennas.[1,2,3] Other practical effects which reduce the main beam gain are relatively easy to analyze from standard array analysis techniques which use vector addition of the elements in the far field. The results of quantization effects for multiple beam antennas with interaction between beams is dependent on the steering algorithm used. For these antennas, Monte Carlo methods offer an efficient technique for evaluation the effects of quantization and even optimizing the steering algorithm to minimize these effects.

Random errors in components used for phased array fabrication are usually governed by allocating the maximum phase and amplitude deviation at the element based on the desired maximum root-mean-square (RMS) sidelobe level.[3,4,5] The individual contributions of the components are combined in root-sum-squared (RSS) to obtain the element deviation. However, this type of analysis assumes independent errors in the components. Here again, Monte Carlo analysis allows a detailed understanding of the relationship between these types of errors and the resulting array performance.

2. MONTE CARLO METHOD

The Monte Carlo method, named after the gaming casinos in the namesake country, is a procedure where a statistical problem, a deterministic problem of many variables, or a combination is solved by running a large number of trials (simulations) and assessing the performance results in probabilistic fashion. One early paper used this method to verify analytical results of beam pointing error in phased arrays; the computer used was the famous IBM 360![6] The Monte Carlo method is becoming much more practical and effective as the power of desktop computers increases. Since the effectiveness of the results increases with the number of trials (roughly as the square root of N), a large number of simulations must be performed for accurate results, especially at the extremes of the probability distribution curves. The best efficiency is obtained by using smaller numbers of trials to determine gross effects, and once these effects have been mitigated, using a larger number of trials to assess the final impact of the architecture on system performance.

3. MULTIPLE BEAM QUANTIZATION LOSS AND ALGORITHM OPTIMIZATION

The first example of the benefits of the Monte Carlo method is illustrated by a multiple beam transmit phased array. The requirement for this antenna is to transmit a single signal to several users with the EIRP at each beam determined by the path loss to the target. Since the quantization effects are deterministic, but the number of variables (beam positions, number of phase and amplitude bits, algorithm, etc.) is large, the Monte Carlo method was used. Figure 1 shows a sample radiation distribution plot (RDP) for a 10 beam case. By comparing the probability distribution curves for the reference excitation (no quantization) of a random set of beam positions (repeated for each trial) with the shifted curves that result when the amplitude and phase is quantized, the effect of quantization can be determined. As a verification of the approach and a check of the accuracy of the coding, a single beam case was modeled and compared to the values given in Skolnik for phase quantization.[3] The single beam phase quantization effect is shown in Figure 2. Figures 3 and 4 show the multiple beam phase and amplitude quantization effects for the same array with 1024 elements given the initial beamsteering algorithm.

The beamsteering algorithm must consider the path loss to the user and position of the user in setting the phase and amplitude weights in the array. Interaction between beams can cause a change in the desired gain. The Monte Carlo method offers a means of investigating the cause and effect relationship between algorithm and resulting gain change. The initial algorithm assumes relative independence between beams and included a margin to account for the remaining interaction. The curve in Figure 5 shows a low probability, but rather large change in gain (extra loss). Investigation of this effect is easy because the random beam positions are stored and the effect can be correlated to a given case. Further analysis revealed the potential for significant improvement in the algorithm through consideration of one type of interaction. Figure 6 shows the benefit of this new beamsteering method and the lower required margin to offset the remaining interaction.

4. CHARACTERIZATION OF RANDOM ERRORS IN ARRAY CONSTRUCTION

Phased array antennas are constructed of elements, phase shifters, power dividers, and for active arrays, amplifiers. Like all fabrication operations, these components very from unit to unit in parameters such as phase, amplitude, compression point, group delay, etc. The Monte Carlo method offers a practical means of analyzing the effects of these variations and conversely specifying the limits on the variations. One advantage of Monte Carlo techniques over the typical RSS combining of errors in the element chain is that the actual probability density function for each component can be used. This can be especially useful when analyzing temperature gradients in arrays, as the mean and shape of the probability density function changes with temperature.

The Monte Carlo method was used to analyze the random phase and amplitude effects on performance for a spaceborne radiometer antenna. The performance of the antenna shown in Figure 7 at a 42 degree scan angle was evaluated for various random excitation amplitude and phase errors associated with the fabrication and component tolerances. The results of 200 Monte Carlo runs are shown in Table 1. Three bit phase shifters were used for all cases in this analysis. Based on the results in Skolnik[3], the use of 3-bit phase shifters will result in approximately 0.23 dB loss in directivity over the ideal phase case. The first entry in the chart shows the Monte Carlo simulation result of 0.21 dB. Random phase and amplitude errors were added to the 3 bit quantized phase distribution. This data provides a concise means of characterizing the antenna so that the system performance of the instrument can be analyzed.

Table 1. Statistical data for the radiometer random error effects are summarized.

CASE	**EXCITATION ERROR PARAMETERS**	**DIRECTIVITY LOSS**	**PEAK SIDELOBE LEVEL**	**PEAK SIDELOBE LOCATION**	**PEAK OF BEAM LOCATION**	**BEAMWIDTH**
	STD OF PHS/AMP	STD/MEAN	STD/MEAN	NOMINAL	STD/MEAN	STD/MEAN
	(deg./amp.)	(dB)	(dB)	(degrees)	(degrees)	(degrees)
1	0/0	.00/-.21	.00/-12.4	39.7	0/42	0/1.44
2	5/0	.02/-.25	.17/-12.4	39.6,39.8,44.4	0/42	0/1.44
3	10/0	.04/-.35	.35/-12.2	39.7,44.4	0/42	0/1.44
4	0/.05	.05/-.22	.14/-12.4	39.7	0/42	0/1.44
5	0/.1	.09/-.21	.28/-12.4	39.7	0/42	0/1.44
6	5/.1	.11/-.25	.35/-12.4	39.6,39.8,44.4	0/42	0/1.44
7	10/.1	.11/-.34	.45/-12.2	39.7,44.4	0/42	0/1.44

note: The quantization directivity loss has not been taken out which is approximately .21 dB.

5. BEAMSTEERING OPTIMIZATION OF A SPHERICAL ARRAY

The operation of the Electronically Steered Spherical Array (ESSA) has been previously described.[7,8] This array uses cluster switching of elements on the sphere to alter the beam pointing direction. The primary steering algorithm uses the N elements with the smallest dot product between the desired pointing angle and the element location. The number of elements has been determined by analyzing a uniformly illuminated sphere and determining the radius at maximum gain. However, the element cluster only has a circular shape for 6 or 7 elements at the first increment greater than one element depending on whether the center element defines a pentagonal or hexagonal lattice. The Monte Carlo method provides a means of monitoring the gain versus number of elements in different regions of the coverage envelope. Figure 8 shows the coverage reduction for the ESSA antenna when the number of elements used varies from 6 to 9. The optimum number of elements in the beam is 7 based on these Monte Carlo cases.

6. CONCLUSION

The examples in this paper show many benefits of the Monte Carlo method in phased array analysis. With current computing resources typically available to the antenna engineer, this technique often characterizes the practical effects of both deterministic and random variations in phased array parameters with less human effort than other approaches. The method is most applicable to problems involving complex interactions of deterministic and/or random variables.

REFERENCES

1. Miller, C. J. (1964) Minimizing the Effects of Phase Quantization Errors in an Electronically Scanned Array, Proc. 1964 Symposium on Electronically Scanned Array Techniques and Applications, RADC-TDR-64-225, vol. 1, pp. 17-38, RADC, Griffiss Air Force Base, NY. ADA448481

2. Brown, J., unpublished communication, 1951.

3. Skolnik, M., (1990) Radar Handbook, McGraw Hill, pp. 7.39-7.49.

4. Ruze, J. (1952) Physical Limitations on Antennas, MIT Res. Lab. Electron. Tech. Rept. 248.

5. Allen, J. L. (1963) The Theory of Array Antennas, MIT Lincoln Lab. Rept. 323.

6. Carver, K. R., Cooper, W. K., and Stutzman, W. L. (1971) Planar Phased Array Beam-Pointing Errors, Southwestern I.E.E.E. Conference and Exhibition, Las Cruces, NM.

7. Kudrna, K. and Hockensmith, R. P., (1982) The Electronic Switching Spherical Array (ESSA) Antenna for the Earth Radiation Budget Spacecraft (ERBS), International Telemetering Conference Proceedings, San Diego, CA, vol. XVIII, pp. 271-288.

8. Taylor, T. H. Jr., (1982) Electronically Steerable Spherical Array Provides Multiple Beam Coverage of Hemisphere, Antenna Applications Symposium, Allerton, IL.

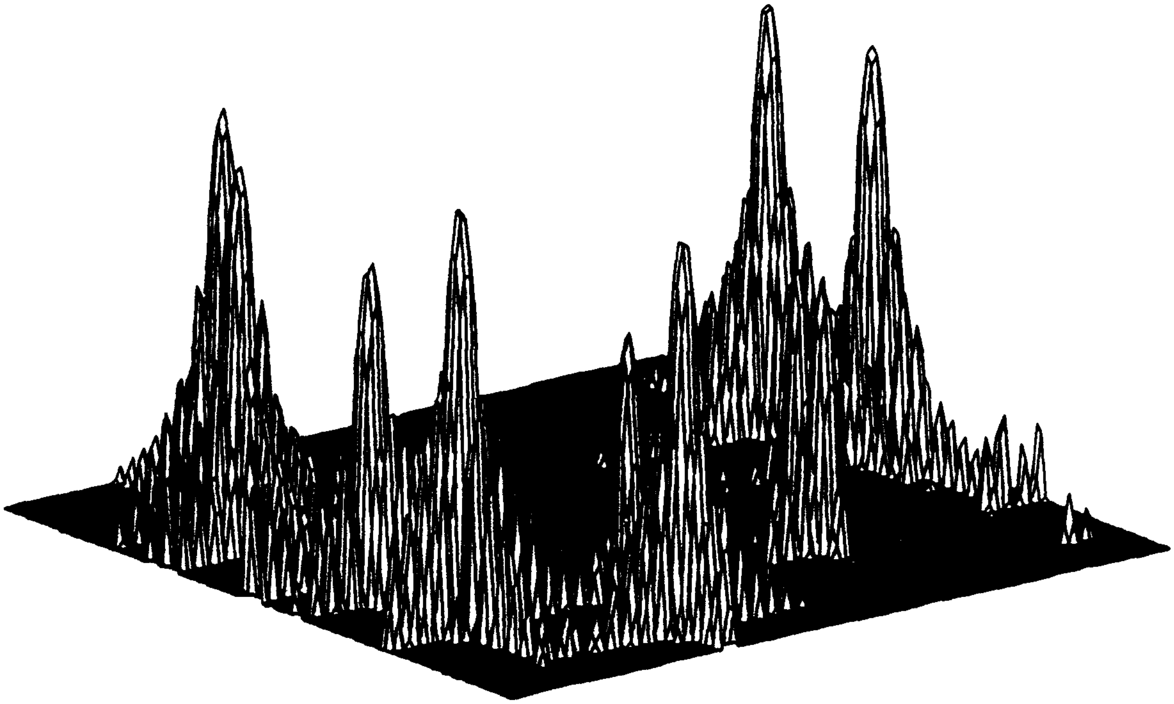


Figure 1a. Plot shows one case from the Monte Carlo analysis of quantization effects.

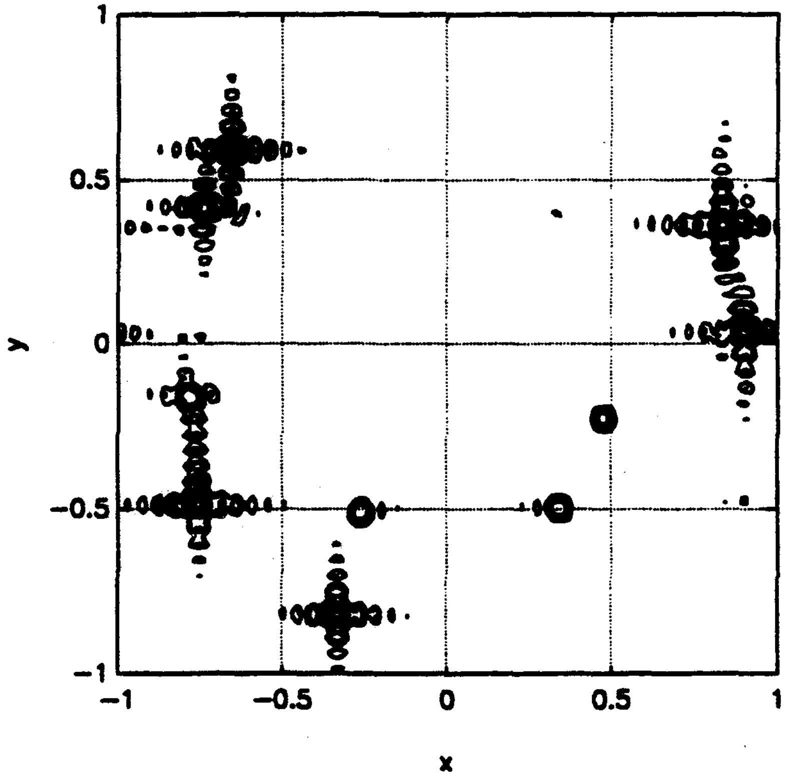


Figure 1b. Radiation distribution contour plot shows locations for one 10 beam case.

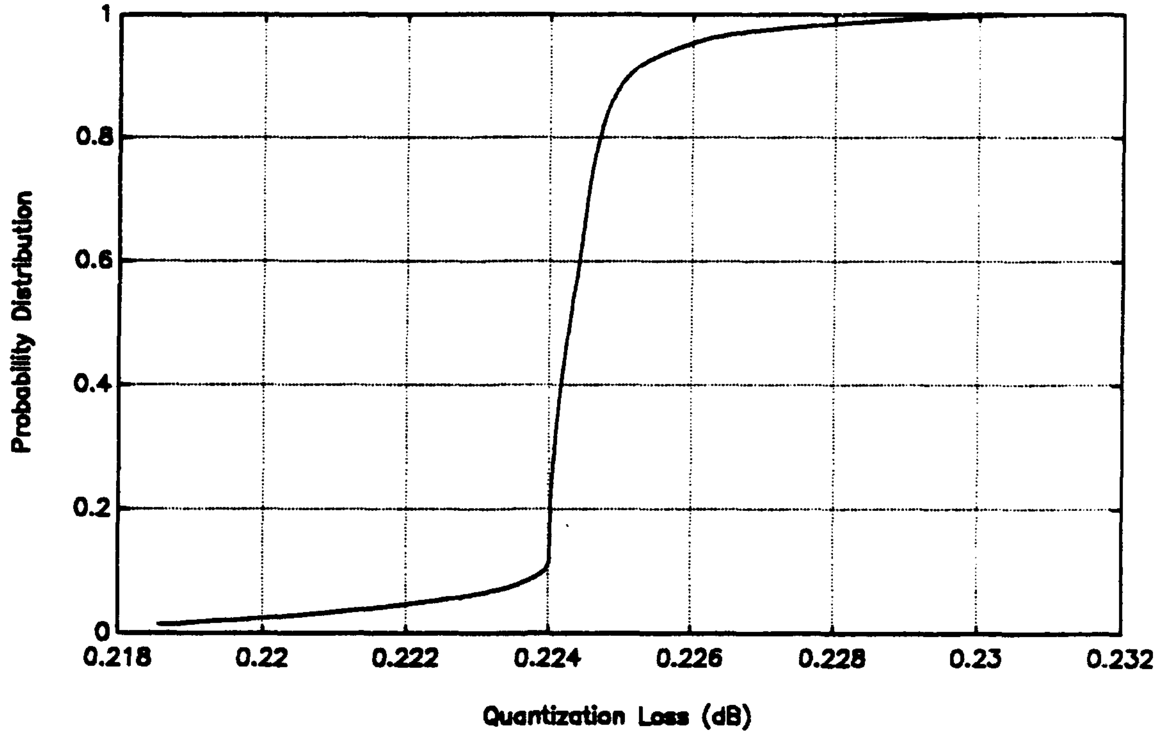


Figure 2. Single beam phase quantization for 3 bits agrees with closed form results.

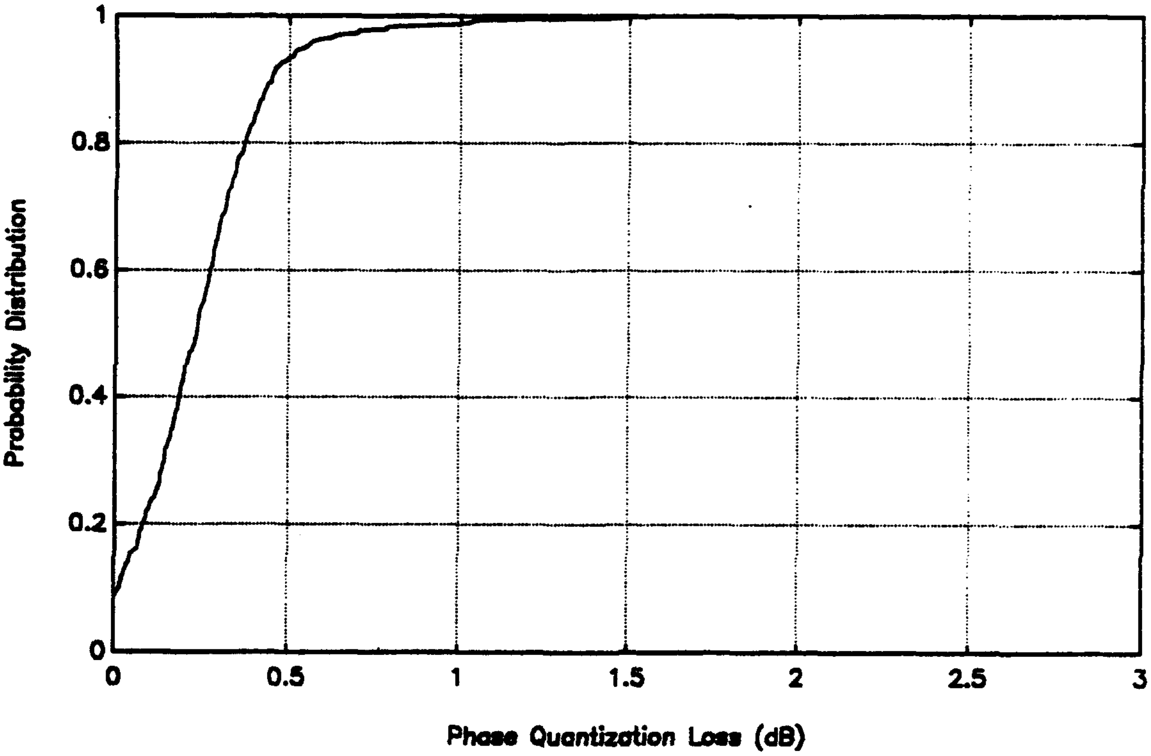


Figure 3. Phase quantization in 10 beam array shows larger variation due to interactions.

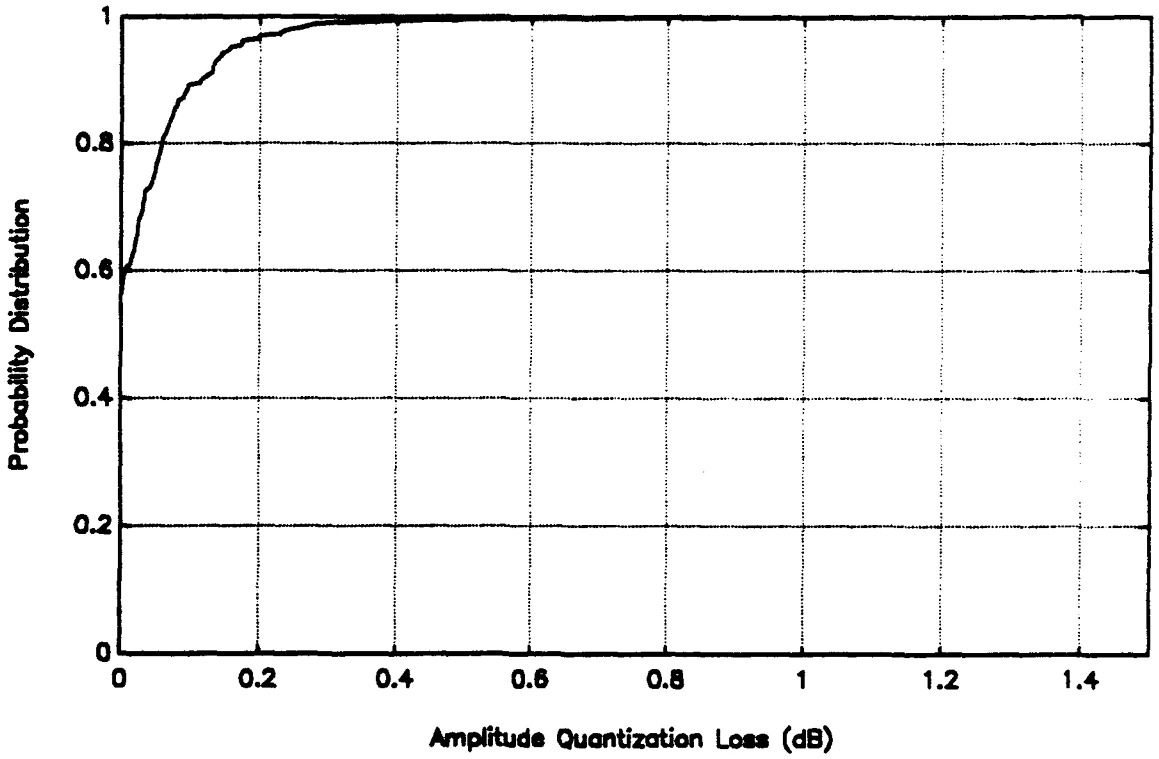


Figure 4. Three bit amplitude quantization loss has less impact than phase for 10 beams.

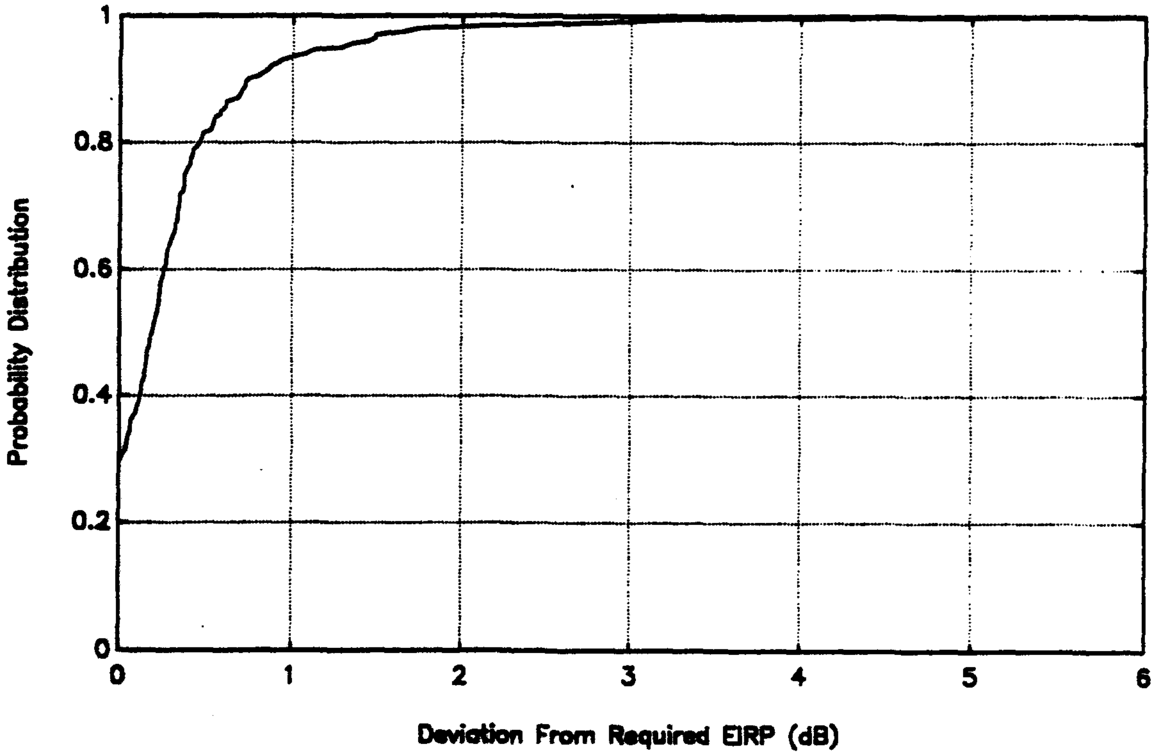


Figure 5. Original beamsteering algorithm shows a few large gain degradation cases.

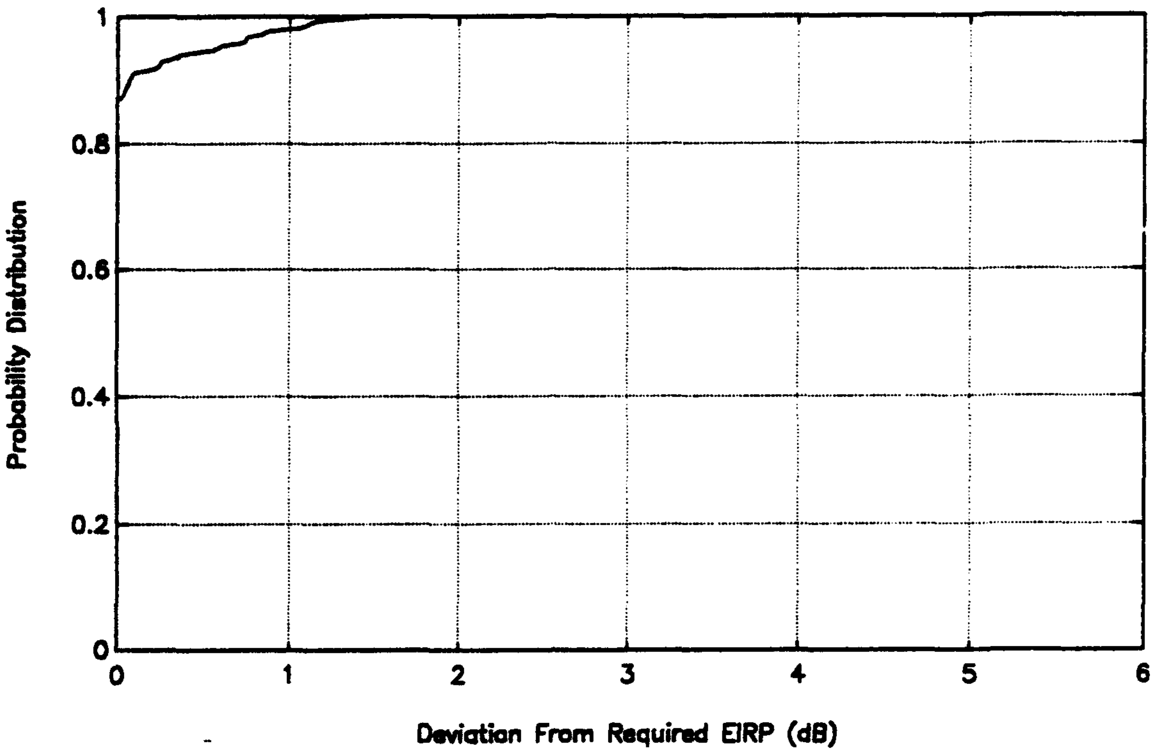


Figure 6. Modified beam steering algorithm minimizes gain degradation.

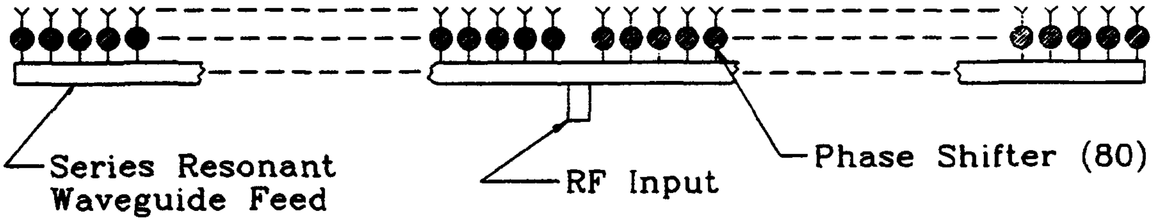

Series Resonant
Waveguide Feed
RF Input
Phase Shifter (80)

Figure 7. This 80 array model is used to evaluate random phase and amplitude errors.

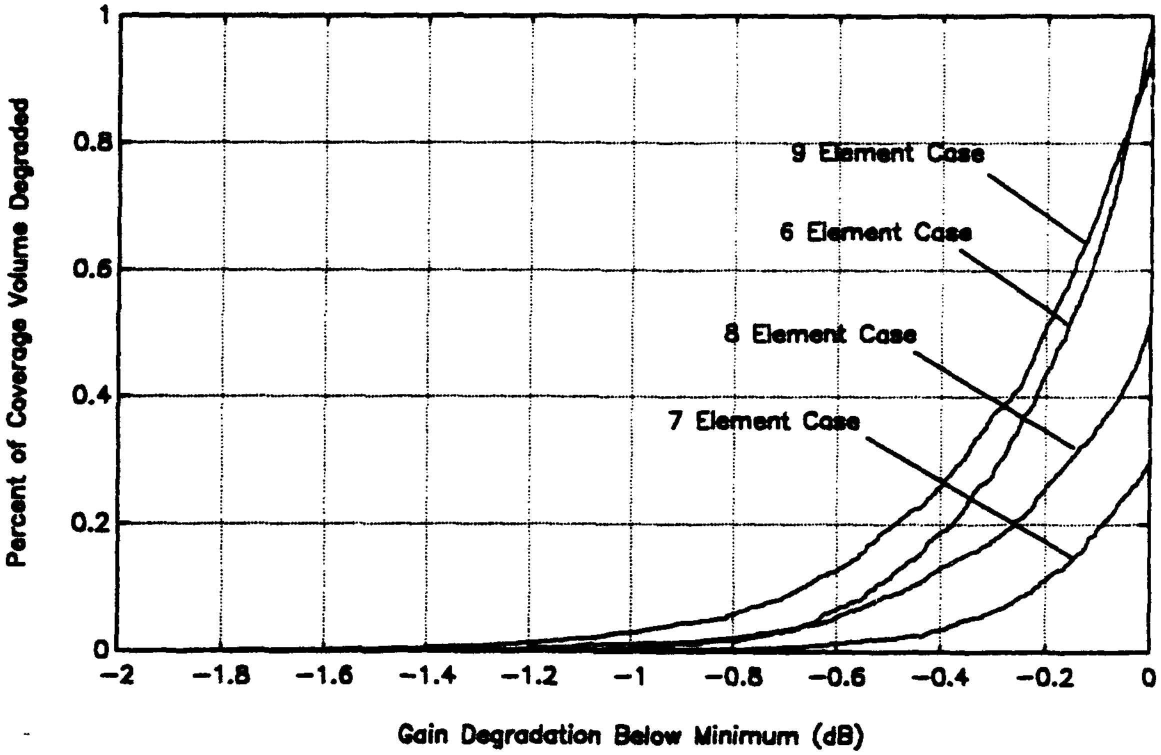


Figure 8. Number of elements per beam including errors is optimized in spherical array.

DIGITAL BEAMFORMING

Warren F. Brandow IV, Lt. USAF
William R. Humbert

ABSTRACT

Last year we detailed the engineering difficulties involved in building a digital beamformer. This year we will demonstrate two reasons for using a digital beamformer. First, the flexibility provided by a digital beamforming system will be demonstrated by providing the results of implementing different nonlinear algorithms such as adaptive nulling, super resolution, and array correction. Second, the accuracy of such a system will be demonstrated by error analysis and correction. Flexibility and accuracy are two very desirable attributes of today's radar and communication systems, and with the cost of digital beamforming system hardware on the decline, such hardware should be affordable in the near future.

I. FLEXIBILITY

Figure 1 shows a block diagram of a phased array antenna. In an analog beamformer the data is not easily accessible until after the beam is formed. Because information is digitized before the beam is formed in a digital beamformer, data is accessible both before and after the beam is formed. This enables the user to, in effect, sample and gain information about the environment. This brings the antenna engineer, unfortunately, into the arena of array signal processing. As a distinguished antenna engineer whom we met at Allerton last year put it, "These days they just stick up a dipole and process the hell out of it!"

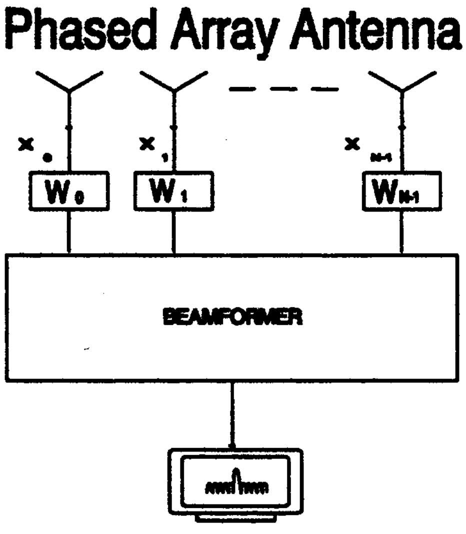


FIGURE 1

There are many nonlinear, well understood algorithms that are applicable to digital beamforming. We will demonstrate two; sample matrix inversion (adaptive nulling), and the MUSIC algorithm (super resolution).

It is not our intention to discuss the theory of these algorithms, but their practical implementation.

First we will demonstrate sample matrix inversion. This algorithm involves sampling the environment and building a covariance matrix.

$$C=\frac{1}{N}\sum_{j=1}^{N}X_jX_j^{T} \qquad (1)$$

where,

C= the covariance matrix
N= the number of time samples
j= the time index
T= complex conjugate transpose
X= the signal vector

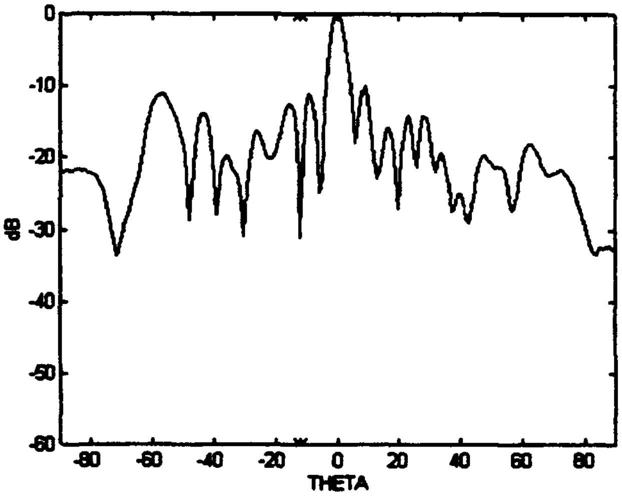
Once the covariance matrix is obtained, the weight vector which will null unwanted jammer signals is given by:

$$W=C^{-1}\bullet S^{*} \qquad (2)$$

where,

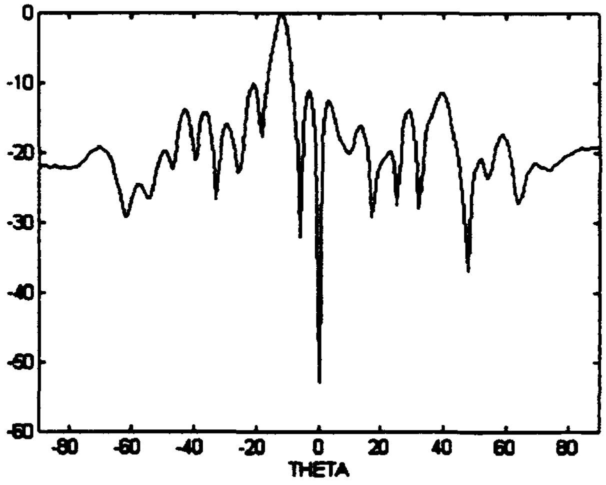
W= the weight vector that will null jammer signals
C= the covariance matrix
S= the weight vector that points in the direction of the
desired signal

FIGURE 2



The object of adaptive pattern nulling is to place nulls in the direction of jammers. To test this algorithm we placed a jammer at -12 degrees. Because the we were testing Applebaum's algorithm, the environment was sampled with the desired signal off. After running the SMI algorithm, we turned the jammer off, the desired signal on, and took an antenna pattern. Figure 2 is what it looked like. As you can see, there is a null in the direction of the jammer, but it is not as deep as we expected it to be. This result seemed to imply that we were not completely nulling the jammer, so we measured the actual power of the jammer after the beam was formed, before and after adapting. That measurement indicated the jammer signal was attenuated 56 dB, or the full dynamic range of the antenna.

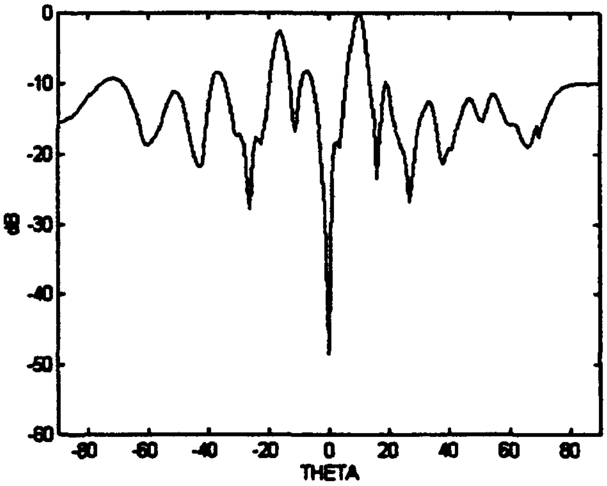
FIGURE 3



This apparent discrepancy was resolved by Hans Steyskal. He pointed out that our jammer signal does not consist of simply a plane wave, the way it does in computer simulations, but multipath from the jammer as well. The SMI algorithm nulls out the total received jammer signal. When we took the antenna pattern with the desired signal on, the illumination came from a different direction and hence the total received signal, including multipath, was diffemt. Therefore it was not nulled entirely. To verify this, we concluded that if we measured the antenna pattern with the jammer on and the desired source off, we should get the nice deep null in the direction of the desired signal. Hence, Figure 3. Figure 4 is the same, but with two jammers.

One handy feature of a digital beamformer is the ability to change the weight vector almost instantly. This makes it easy to apply or change an amplitude taper. When we tried to adapt to jammers and use an amplitude taper, the taper didn't work. This is because the noise eigenvalues of the covariance matrix are not equal, as they are in most computer simulations. To correct this problem, we diagonally loaded the covariance matrix:

FIGURE 4



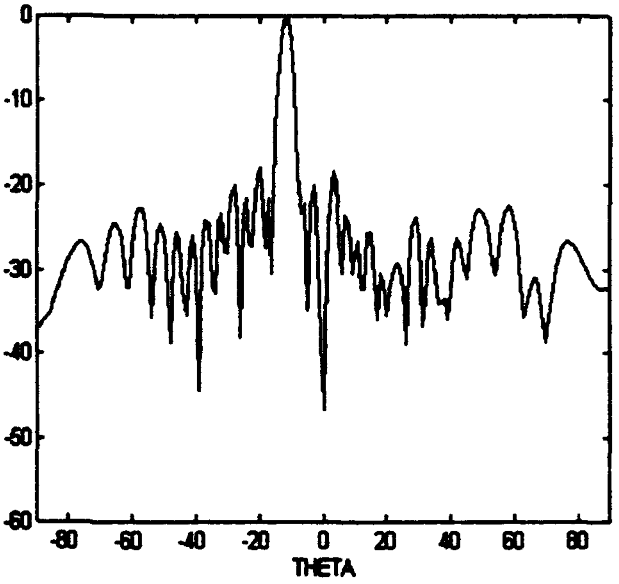
$$C = C + 100 \bullet I \tag{3}$$

where,

C= the covariance matrix
I= the identity matrix

Figure 5 shows that this approach was successful (20 dB Taylor). However, we all know that you don't get something for nothing, and the price paid for good sidelobe structure is about 3 dB of null depth. At first glance this doesn't seem like much, but it can be significant.

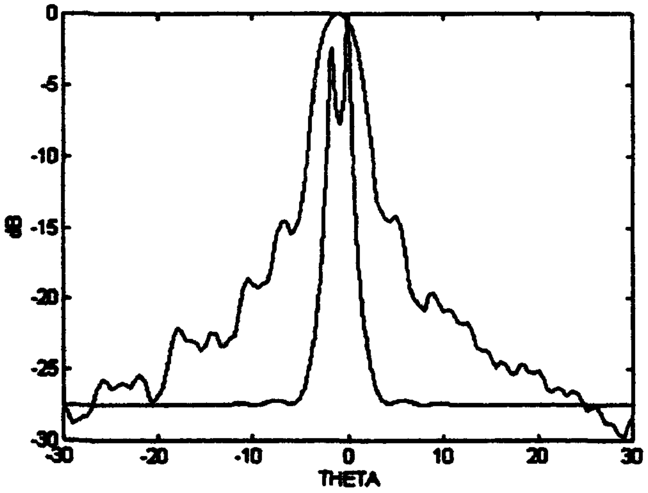
FIGURE 5



With our system we have the ability to Amplitude modulate the sources, and actually listen to the real time output of the beamformer. We can do this because we have a D/A converter at the output of the beamformer which feeds a speaker.

When we don't use diagonal loading and listen to the beamformer output, we can't hear the jammer at all after adapting. The jammer has been attenuated down to the thermal noise level. When we do use diagonal loading to get a good sidelobe level and listen to the output of the beamformer, we can still hear the jammer clearly (even though it is still attenuated severely) because it is 3 dB above the thermal noise.

FIGURE 6



The MUSIC algorithm is like SMI in that it also samples the environment and builds a covariance matrix. It then uses the information in the covariance matrix to resolve signals that are closer than a beam width apart.

FIGURE 7

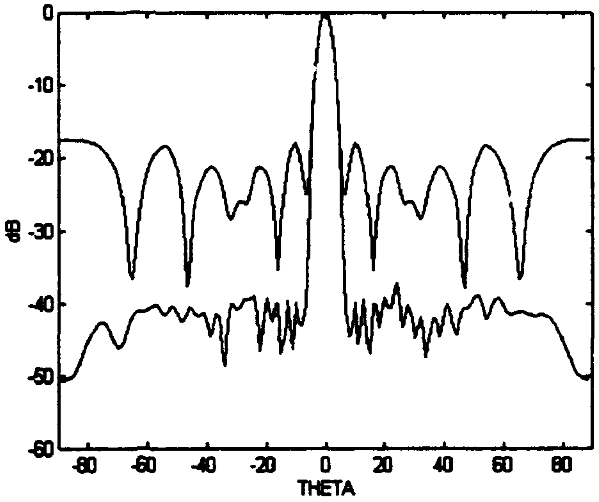


Figure 6 shows two signals separated by less than half a beam width (1.7 degrees). It is easy to see that MUSIC is superior to simply scanning the antenna.

Once you have information about the signals incident on the antenna, you can use that information for other algorithms, like Dr. Mailloux's Antenna Correction. This algorithm uses the information MUSIC provides to digitally reconstruct element signals that are lost because antenna elements are damaged. This is possible because MUSIC will still work with damaged antenna elements. Figure 7 shows an antenna pattern with 3 elements damaged, before and after Dr. Mailloux's Antenna Correction.

Sampling the environment and adapting to it is a clear example of the flexibility of digital beamforming. We have demonstrated that it works not just in computers, but with real data as well. Furthermore, the theory is well understood (by array signal processing engineers) and the algorithms are easy to implement.

II. ACCURACY

FIGURE 8

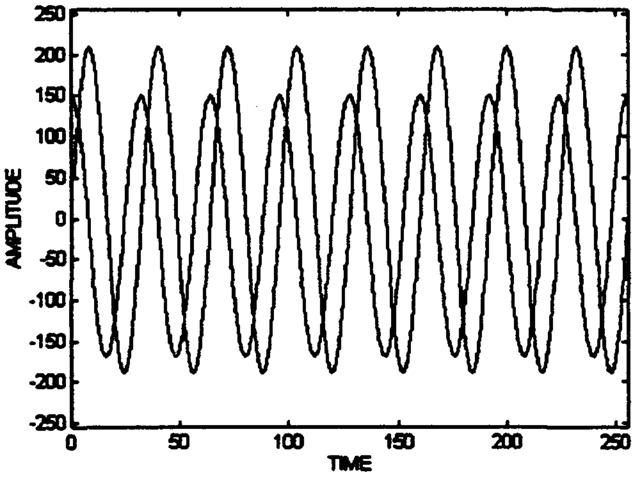


Figure 8 is an example of a digitally sampled time signal. Both inphase and quadrature components are shown. Ideally both components would be identical (except for a 90 degree phase shift) and centered around zero. In reality they are not. Mathematically, the signal should look like:

$$S = A \bullet \sin(\omega t) + A \bullet \cos(\omega t) \bullet i \qquad (4)$$

But in reality it is:

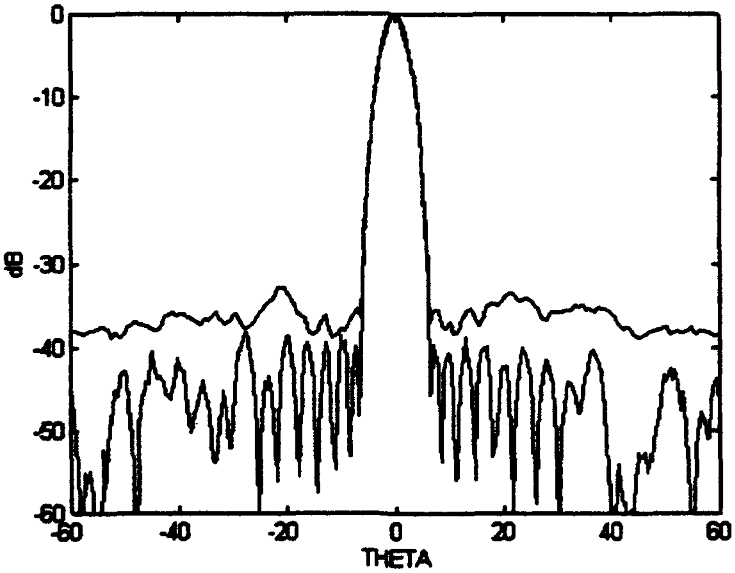
$$S = [A \bullet \sin(\omega t) + 10] + [.8 \bullet A \bullet \cos(\omega t + 2^{\circ}) - 10] \bullet i \qquad (5)$$

In addition to the DC offset error, the Q portion is not exactly 90 degrees out of phase nor is it's amplitude equal to the I portion amplitude. These errors are all introduced before the signal is digitized. Once the signal is digitized, no further errors are introduced (except for dynamic range limitations).

Knowing what these errors were, we wanted to know if it was worth the time and money to replace hardware and correct the errors. Lars Petterson suggested that we use the FFT to filter out the error effects. By setting the baseband frequency to 15.625 Khz, we forced the errors to show up at different frequencies than the desired signal in the frequency spectrum. We then zeroed those frequencies corresponding to errors and performed an IFFT. Figure 10 is a comparison of antenna patters with and without the errors. Our conclusion is that correcting these problems would be money well spent.

Because our beamformer is a 9 bit machine, what we have is a 32 element antenna with accurate and reliable 9 bit phase shifters. This is old technology. We could do much better with today's technology (12 bits). If you combine the ability to correct errors generated on the analog side of the hardware, with high accuracy phase shifters, and no error introduction on the digital side of the hardware, you can see the potential for a very precise antenna system. We have demonstrated that such precision is achievable in a real digital beamformer both by simply getting the hardware to work, and by error correction.

FIGURE 9



HIGH TEMPERATURE SUPERCONDUCTING ANTENNA ARRAY DEVELOPMENT

Livio D. Poles, James P. Kenney, Edward Martin, Jeffrey S. Herd

Rome Laboratory, Electromagnetics and Reliability Directorate
Hanscom, A.F.B. Massachusetts

The development of high-temperature superconducting films has sparked much interest within the phased array antenna community. Using superconducting films for the RF power distribution network of large array antennas can considerably reduce ohmic losses and hence increase antenna efficiency. In this paper we will discuss the results of our design process adopted to produce a 12 GHz, sixteen element (4X4), high- temperature Yttrium Barium Copper Oxide (YBCO) superconducting array.

1. INTRODUCTION

Researchers have proposed [1,2,3] using superconducting strip lines for the power distribution networks of large array antennas in order to considerably enhance the array efficiency. Lower operating temperatures enhance the array antenna noise figure. These benefits are very important for space borne antenna systems where RF power is at a premium.

2. ARRAY ANTENNA ARCHITECTURE

A novel multi-layer 4X4 element microstrip array [4] with the upper patches at room temperature separated and thermally isolated from the superconducting lower patches is used as the basic antenna design architecture. Figure 1 shows the architectural design for the antenna array. The stacked patch array consists of three dielectric layers which are, from top to bottom, a quartz superstrate, a vacuum gap and a Lanthanum Aluminate substrate. Silver radiating patches were deposited on the top surface of the quartz window. A lower network of power dividers and microstrip patches was patterned on the top surface of the $LaAlO_3$ substrate from a thin film of high Tc superconductor (YBCO), while a silver layer deposited on the bottom provided the ground plane.

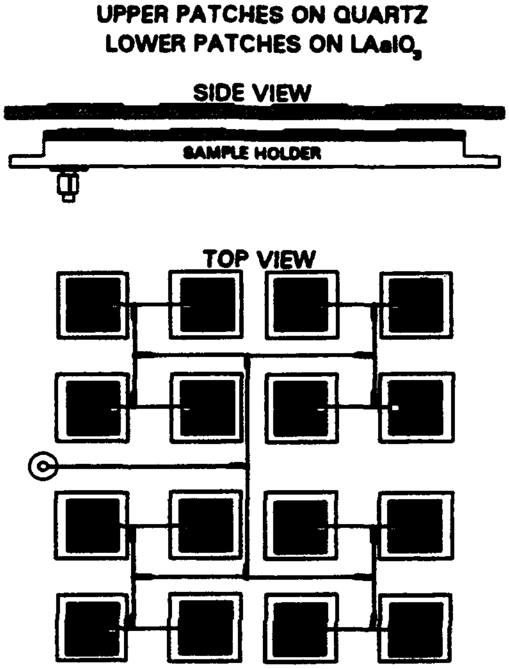


Fig. 1 MICROSTRIP ARRAY

The upper patches are exited by capacitive coupling to the lower patches, and thus broadens the bandwidth by a factor of five over that of a single patch. A multi-layer spectral domain Green's function was used with the method of moments to design the microstrip feed network and antennas [5]. This approach models all mutual coupling effects, including surface wave resonance's. For simplicity, the analysis assumes that the superconductor has infinite conductivity. This appears to have been sufficient for the modeling of impedance and radiation properties of the microstrip antennas and the modal phase velocity and characteristic impedance of the microstrip lines. Several 4X4 arrays were designed and fabricated, with the lower layer feed networks fabricated in both YBCO and silver for comparison.

3. CRYO-COOLER

The multi-layer microstrip stacked patch array was mounted inside a Janus continuous transfer, cold finger cryostat, shown in figure 2. The cryostat has two contiguous compartments, one outside the cold finger and one inside. The inside compartment is used to transfer the cryogen, which in turn cools the cold finger.

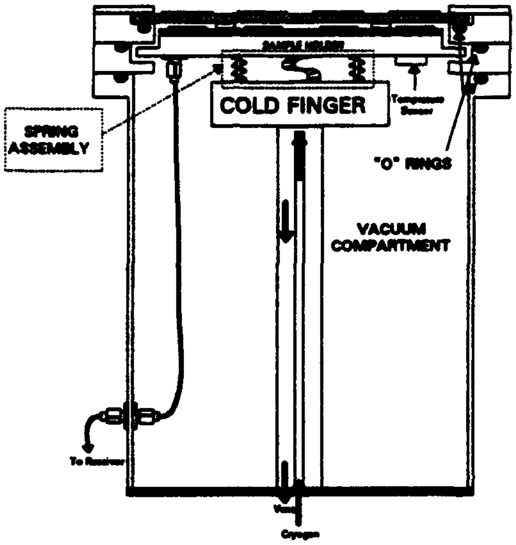


Fig. 2 ARRAY ANTENNA - CRYOSTAT ASSEMBLY

The outer compartment is evacuated to provide thermal isolation. The $LaAlO_3$ substrate with superconducting power divider network and lower patch array is mounted on a brass fixture which is thermally coupled to the cold finger. Liquid helium is transferred from a storage dewar through a flexible transfer line into the inner compartment of the cryostat. The cold finger cools the superconducting array assembly by conduction. The upper patches are virtually at room temperature and are thermally separated from the lower patches by the vacuum gap. This gap assists in maintaining the required thermal isolation between ambient temperatures and superconducting temperatures. For the proximity coupled microstrip antenna architecture it is important to keep the distance between the upper and lower microstrip patches constant. Precisely cut teflon spacers are used to separate the arrays. Since cooling causes the cold finger to shrink in length by 1/32", directly attaching the lower array to the cold finger is not feasible. In our design a specially devised spring assembly provides the compensation for the inherent cold finger shrinkage. When properly designed and constructed this spring assembly keeps the lower array in contact with the spacers and upper array. To attain a good thermal transfer across the spring assembly the lower array is thermally connected to the cold finger by a flexible copper wire braid. A temperature sensor placed on the lower array mounting fixture monitors the lower array temperature during measurement. The lower array power divider network is connected to a standard SMA probe. Wire bonding is used to connect the center pin to the lower patches and power divider network. Cryogenic, stainless steel jacketed, semi rigid coaxial cable (.085- Dia.) provides the RF path between the power divider network and the hermetically sealed SMA to SMA feed-through connector.

4. MEASUREMENT

Antenna radiation patterns and gain/bandwidth measurements were performed with the array at superconducting temperatures. Data acquisition time was minimized by using a fully automated measurement system.

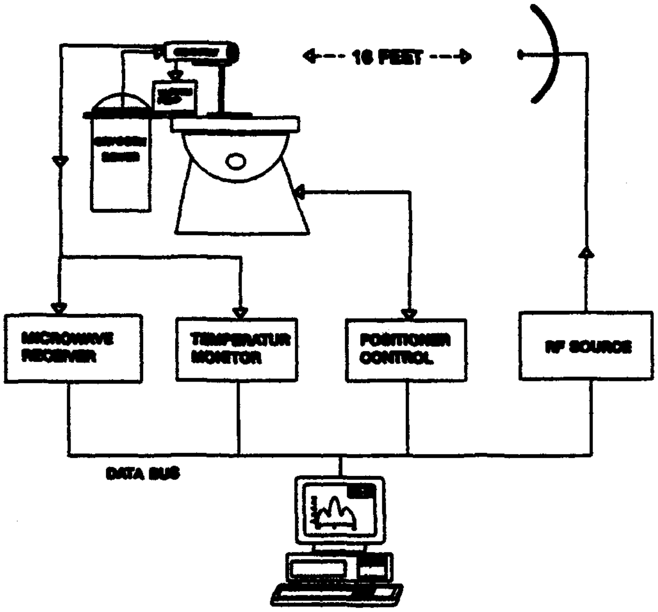


Fig. 3 MEASUREMENT CONFIGURATION

The array "turn-on" and "turn-off" transition temperatures were monitored as a function of time. Figure 3 shows the measurement setup and instrumentation used to perform the data acquisition. The cryostat, which holds the proximity coupled patch array, is mounted horizontally on the antenna positioner. The antenna array is accurately placed over the rotational center of the positioner. The cryogen dewar and cryostat vacuum pump are mounted to the positioner and move with the antenna during measurement. An HP computer controls the SA microwave receiver, SA positioner/controller, HP synthesized source, and reads the array temperature sensor. During the measurement the computer automatically stores the receiver data and the temperature of the antenna. Isotropic gain vs. frequency of the array is measured relative to a standard gain horn. Figure 4 shows that the measured bandwidth for this array geometry is ~10%.

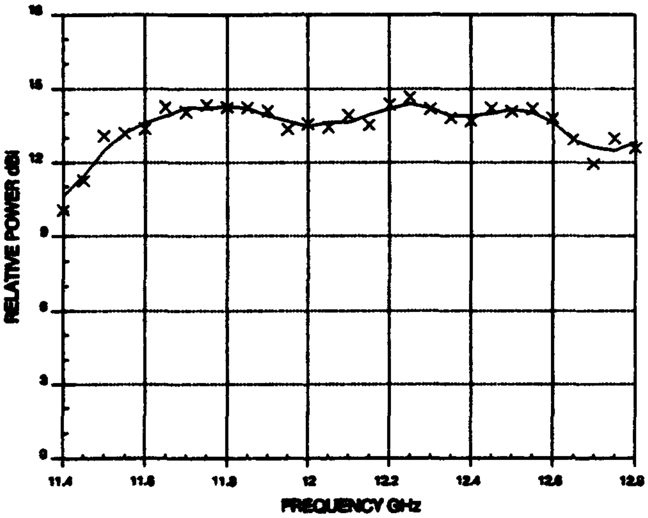


Fig. 4 ISOTROPIC GAIN vs FREQUENCY

The transition temperature of the array is measured vs. time by monitoring the received power while recording the array temperature at one minute intervals. Figure 5 shows the array "turn-on" and "turn-off" temperatures as a function of time. The steep transition between the "turn-on" and "turn-off" temperatures indicates, in part, a good quality superconducting film..

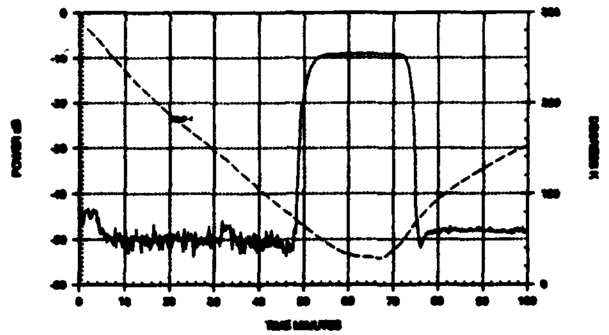


Fig. 5 TRANSITION TEMPERATURE

A well behaved antenna radiation pattern, taken at the center frequency of 12.2 GHz, is shown in figure 6. The measured isotropic gain for this array is 14.8 dBi and compares well with predicted results.

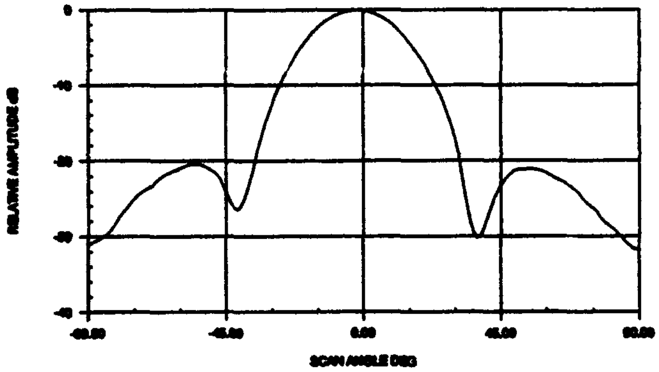


Fig. 6 H-PLANE RADIATION PATTERN

4. CONCLUSION

Experimental measurements of the far field radiation patterns show excellent agreement with predicted beamwidth, directive gain, resonant frequency and bandwidth for the broadside and scanned beams. Integrating the proximity coupled array antenna architecture as part of the cryostat has eliminated the need for using a radome. This antenna-cryostat implementation technique coupled with superior antenna performance make this emerging antenna technology very attractive for many space borne antenna applications.

REFERENCES

[1]. M.A. Richard, K.B. Bhasin, C. Gilbert, S. Metzler, and P.C. Clapsy "Measurement Techniques for Cryogenic KA-Band Microstrip Antennas," AMTA Proceedings 1991.

[2]. R.C. Hansen, "Superconducting Antennas," IEEE Trans on Aerospace and Electronic Systems, vol. 26, no.2 pp.345-354, March 1990.

[3]. J.S. Herd, D. Hayes, J.P. Kenney, L. Poles, K.G. Herd, W.G. Lyons, 'Experimental Results on a Scanned Beam Microstrip Antenna Array With a Proximity Coupled YBCO Feed Network", IEEE Trans. on Applied Superconductivity, Vol.3,No 1, pp. 2840-2843, March 1993.

[4]. J.S. Herd , 'Scanning Impedance of Proximity Coupled Rectangular Microstrip Antenna Arrays", Ph.D Thesis, University of Massachusetts, 1989.

[5]. J.S. Herd , "Full Wave Analysis of Proximity Coupled Rectangular Microstrip Antenna Arrays", Electromagnetics, Jan 1992.

DEVELOPMENT OF A DUAL FREQUENCY ARRAY FOR ACTS SMALL TERMINAL

Martin L. Zimmerman

Analex Corporation

Richard Q. Lee

NASA Lewis Research Center

Abstract

The Advanced Communications Technology Satellite (ACTS) is scheduled to be launched in the summer of 1993. Small (less than 1 m in diameter), ground-based terminals will be used for conducting experiments involving this satellite. The Lewis Research Center is developing a passive array that could serve as a small terminal.

This array would transmit linearly-polarized energy in the 29.0 - 30.0 GHz band, and receive the orthogonal linear polarization in the 19.2 - 20.2 GHz band. The array uses microstrip patches as the radiating element. An FDTD code is used to design the patch elements, as well as the beam forming network. This work is currently in progress. Results will be presented for 4-element and 16-element subarrays.

1. Introduction

The Advanced Communications Technology Satellite (ACTS) is scheduled to be launched in the summer of 1993. The purpose of this satellite will be to demonstrate communications technology in the Ka-band. A number of experiments are planned with ground users. Some of these experiments will require a small, easily-transportable ground station that will operate from a fixed position. A ground station using a small (0.6 or 0.8 m) reflector antenna designed by Prodelin Corp. for a NASA Lewis research program is being constructed at the NASA Lewis Research Center. A flat-plate, passive array has the potential to be much smaller in volume (due to the lack of a feed structure) and lighter than this configuration. In addition, use of an array allows the power amplifiers and low noise amplifiers (LNAs) to be distributed. This allows the use of lower power amplifiers (building a 1 W 30 GHz solid-state amplifier is still quite challenging) and makes heat dissipation easier to manage.

2. Requirements

The uplink to the ACTS satellite is at 29.2 GHz, and the downlink is at 19.2 GHz. This represents the largest challenge to an array design, since these frequencies are separated by 50% in bandwidth. Since the object is to replace the 0.6 m reflector antenna, the aperture size of the array should be of the same order of magnitude. With a 1 W power amplifier, the USAT terminal is expected to have 30 dB link margin in clear weather (most of which is needed to compensate for rain fade in foul weather).

3. Design

One design method would be to use separate radiating elements for the two frequencies. This method is being used in an dipole/slot array under development at JPL for their Mobile Communications Program. However, since the uplink and downlink use orthogonal, linear polarizations, a single microstrip patch could be used for both bands, if each set of edges is designed to resonate at one of the pertinent frequencies.

3.1. Feed Network

Initially, a design using one feed network, with two separate feeding points on each patch, was investigated. Unfortunately, a great deal of coupling existed between the two modes, making it difficult to tune the resonant modes independently. The next step was to investigate the use of separate beam-feed-networks (BFNs) for transmit and receive. If designed well, this would give the added advantage of isolation between the receive signal and the much more powerful transmit signal. In the present design, microstrip line BFNs are used for both frequencies. The 30 GHz BFN is on the same surface as the microstrip patches. The 20 GHz BFN is located on the underside of a lower substrate, separated from the patches by a ground plane. The 30 GHz BFN is located on top because the quarter-wave transformers used for matching the 2:1 power dividers are smaller at 30 GHz. The top BFN must share real estate with the radiating patches, and therefore has less space to work with. The feed structures are corporate in nature.

3.2. Feed mechanism

The upper BFN feeds into the microstrip patch using an inset, center-fed microstrip line (Fig. 1). The patch is center-fed in order to reduce coupling to the orthogonal modes of the patch. The lower BFN is coupled to the patches using rectangular slots in the ground plane. The microstrip lines are again centered with respect to the patch in order to reduce coupling to the orthogonal modes of the patch. The slot is offset towards the patch edge in order to achieve a better match. Slot coupling is much easier to manufacture than probe-feeding, especially at Ka-band frequencies. Once the separate substrates are etched, they are aligned and attached with bonding film. The substrates used here are 10 mil RT/Duroid 5880 ($\varepsilon_r = 2.2$) with 3001 bonding film.

Based on work done at JPL[1], adjacent elements are fed out of phase. This is achieved by feeding opposite edges of adjacent patches (Fig. 2).This is compensated for by putting a 180° phase delay on one of each pair of patches. The result should be that the reference-polarization energy for each pair of patches should add, while the cross-polarization energy should cancel out, increasing cross-pol isolation.

4. Results

The design process started with analyzing a single radiating patch. A Finite-Difference Time-Domain (FTDT) code was used for this analysis[2]. FDTD allows for effects such as coupling between patches and slots to be modeled without the delays associated with building and testing antennas repeatedly. Different feeding mechanisms were examined until the design set forth in the previous section was settled upon. Using the data generated by the FDTD code, it is possible to determine the return loss and far-field pattern of the antenna under study. Once a final design was settled on, the patch was fabricated. Return loss was verified using an HP 8510 network analyzer, and far-field patterns were measured in the Lewis far-field test facility.

The design process was then repeated for a 2 x 2 array (4 elements). Two element spacings, 10 x 10 mm and 10 x 12 mm were studied. Once it was determined that the 10 x 10 mm 2 x 2 element array was well-matched, a 4 x 4 (16 element) array was constructed. In this case, due to the large amount of time needed to analyze the antenna using FDTD, the measurements were performed first. At the time of writing, FDTD analysis of 4 x 4 had not been done.

4.1. Single Element

FDTD predicted return loss of -16 dB at 19.2 GHz and -22 dB at 30 GHz. When measured, the return loss was -33 dB at 20 GHz and -36 dB at 30.95 GHz. The difference in resonant frequency between FDTD and measurement is due to discretization errors that result from FDTD quantizing distances into multiples of the grid size.Based on these results, the narrow edge of the patch was increased slightly in order to lower the resonant frequency of the receive band. This slight effect was included in later designs but could be modeled with FDTD without changing the grid size, which was not desired. Using a near-field to far-field transform, directivities of 4.1 dB at 19.2 GHz, and 4.4 dB at 30 GHz were predicted. Measured patterns were in agreement with predicted results.

4.2. 2 x 2 Arrays

Two 2 x 2 designs were tested. One design used a 10 x 12 mm spacing and the other used a 10 x 10 mm spacing. Both designs worked well, but since the 10 x 10 mm spacing was chosen for the 4 x 4 array, only results for this array are shown.

For the receive band, FDTD predicted a return loss of -25 dB at 19.1 GHz with a bandwidth (return loss less than -10 dB) from 18.9 to 19.2 GHz. For the transmit band, FDTD predicted a double null with -28 dB at 30 GHz, and return loss less than -10 dB from 29.4 - 30.5 GHz and 31.6 - 32.1 GHz. The receive band was measured at -18 dB at 20.2 GHz with bandwidth from 19.4 - 20.4 GHz. For the transmit case (for which the patch size had been slightly adjusted), the measured return loss was -36 dB at 29.7 GHz with bandwidth from 29.4 - 30.0 GHz.

The predicted receive far-field pattern had directivity = 9.9 dB at 19.1 GHz (Fig. 3,4). Radiating efficiency (ratio of energy in main feed to forward radiated energy) was 75%. The isolation between the receive and transmit feeds was -47 dB and the front/back radiation ratio was 6.1 dB (due to radiation by the coupling slots). The predicted transmit far-field pattern had direcitvity = 10.6 dB at 30 GHz (Fig. 5,6). Radiating efficiency was 80%. Receive/transmit isolation was -26 dB and the front/back ratio was 13 dB. Figures 3-6 compare the predicted and measured patterns.

4.3. 4 x 4 Array

The 10 x 10 mm spacing was used for the 4 x 4 (16 element) array. At the time of writing, only measured results exist for this array. The return loss for the receive band was less than -10 dB from 16.8 GHz t 19.5 GHz and a frequency sweep in the far-field range showed radiation to be strongest in the 19.2 - 19.9 GHz range. In addition, transmission measurements were taken between the feed points for the receive and transmit BFNs. Transmission was below -30 dB in the frequency bands of interest and averaged about -40 dB. The ref-pol and cross-pol patterns are shown for the phi = 0° cut (Fig. 7) and the phi = 90° cut (Fig. 8). The highest sidelobe was down 11 dB and the cross-pol isolation was 20 dB or greater. Figure 9 shows the 360° pattern for the phi = 0° cut. The front/back ratio is 6 dB. In an actual array based on this design, absorber would be needed to prevent stray signals from the back plane from being received.

The return loss for the transmit band was less than -10 dB from 27.4 GHz to 29.8 GHz and a frequency sweep in the far-field range showed radiation to be strongest in the 28.5 - 30.2 GHz range. The ref-pol and cross-pol patterns are shown for the phi = 0° cut (Fig. 10) and the phi = 90° cut (Fig. 11). *The grating lobes are very noticeable in the phi =* 90° cut (6 dB down) but are down 16 dB in the phi = 0° cut due to the narrower element pattern in this plane. The highest sidelobe is down 9 dB. The cross-pol isolation was at least 18 dB in the main beam region. The 360° far-field pattern is shown in Figure 12 for the phi = 0° cut. The front/back ratio is 25 dB. This means that the absorber necessary for the receive portion of the antenna will not have to absorb much of the power radiated by the transmitting portion of the antenna.

5. Conclusions

A passive subarray for use with the 20/30 GHz communication bands has been demonstrated. The transmit band matches well with the ACTS system, but the receive bandmatching network may need to be adjusted slightly. A 4 x 4 module has been shown to have good far-field patterns in both bands.

Cross-pol isolation was not as good as that demonstrated by Huang[1]. This is probably due to the fact that Huang's antenna was probe fed, which means that the cross-pol energy came completely from the microstrip patches. In this design, there is radiation from the coupling slots and also from the transmit BFN, which is on the same surface as the microstrip patches. This is the probable cause of the higher cross-pol levels predicted and measured. However, it should be noted that construction of a probe-fed Ka-band array would be extremely difficult. Huang's antennas operated at 1.5 GHz.

Grating lobes are noticeable in the phi = 90° cut but are well separated from the main beam. Since it would be difficult to reduce the element spacing any farther, the grating lobe level could only be reduced by increasing the element gain. This could be done with stacked parasitic elements.

Once the 4 x 4 element design has been finalized, the next step will be to construct a 16 x 16 element array. This is the largest such structure that can be fabricated in-house. The eventual goal is to produce an array unit that can be used to replace the reflector antenna presently used in the Lewis mobile terminal test bed (MTTB).

References

1. Huang, J. (1990) Low cross-pol linearly polarized microstrip array, IEEE Symp. Dig.: 1750-1753.
2. Zimmerman, M. L. and Lee, R. Q. (1993) Use of the FDTD method in the design of microstrip antenna arrays, Journal MIMICAE, to appear Oct. 1993.

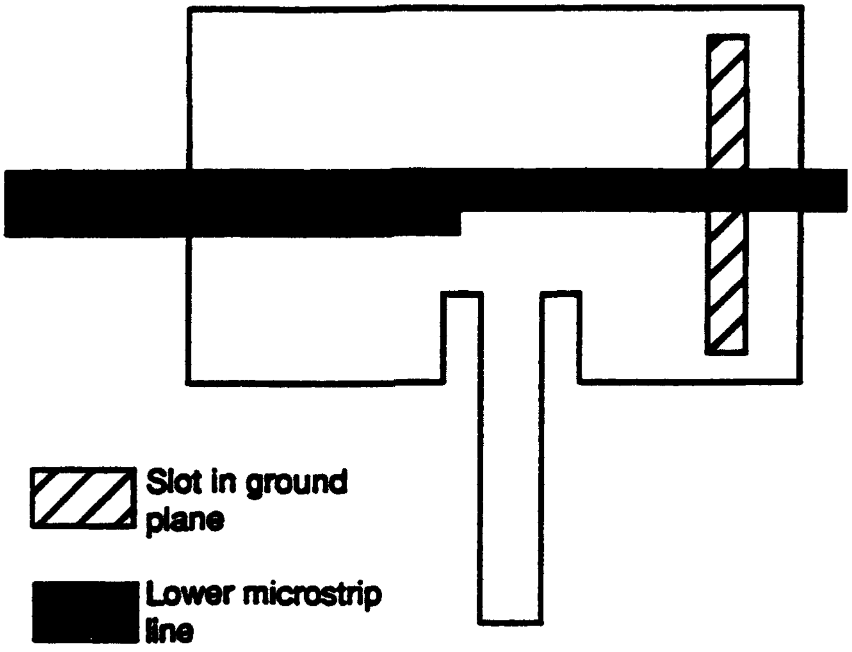

Slot in ground plane
Lower microstrip line

Figure 1. Geometry of 20/30 GHz patch

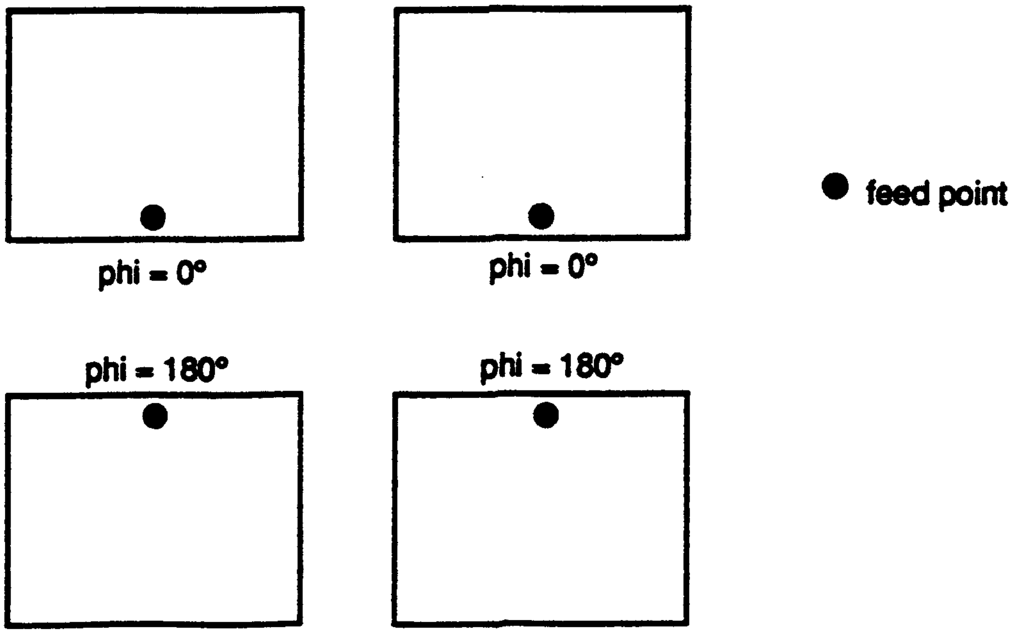

phi = 0°
phi = 0°
phi = 180°
phi = 180°
● feed point

Figure 2. Feeding arrangement for reduction of cross-pol.

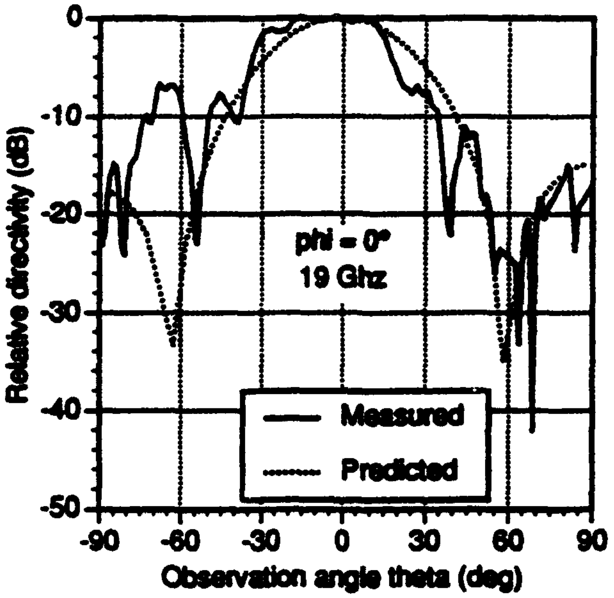


Figure 3. Phi = 0° cut for 2 x 2 array in the receive band.

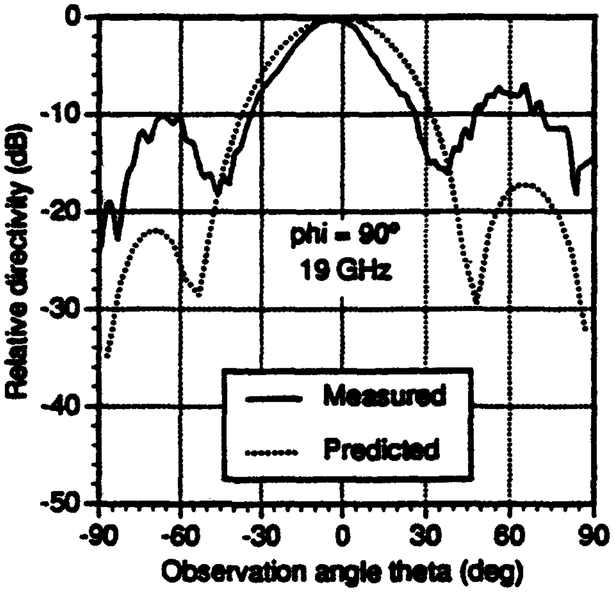


Figure 4. phi =90° cut for 2 x 2 array in the receive band.

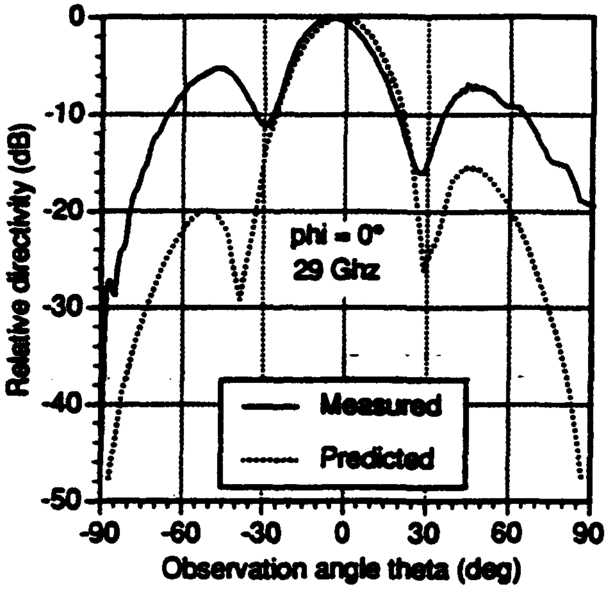


Figure 5. Phi = 0° cut for the 2 x 2 array in the transmit band.

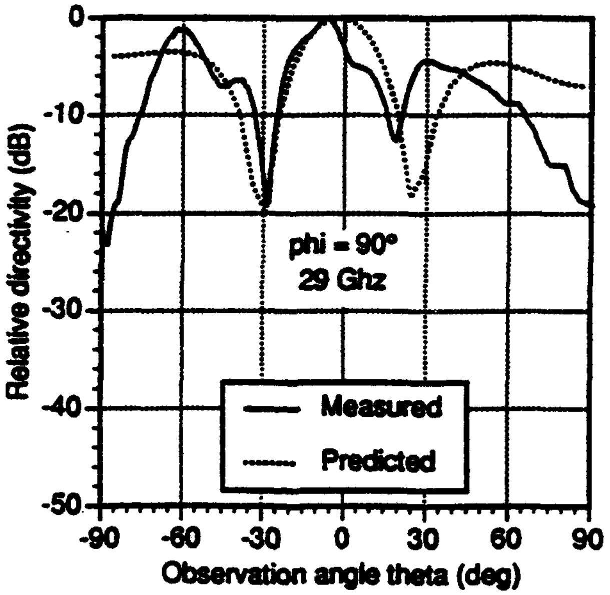


Figure 6. Phi = 90° cut for the 2 x 2 array in the transmit band.

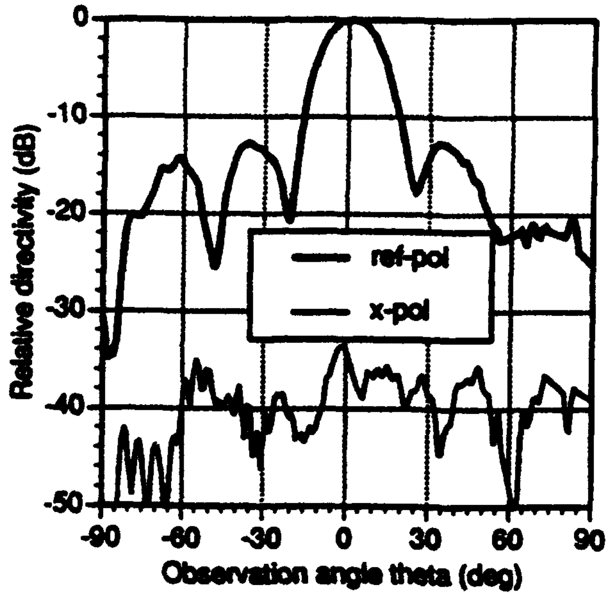


Figure 7. Phi = 0° cut for the 4 x 4 array in the receive band.

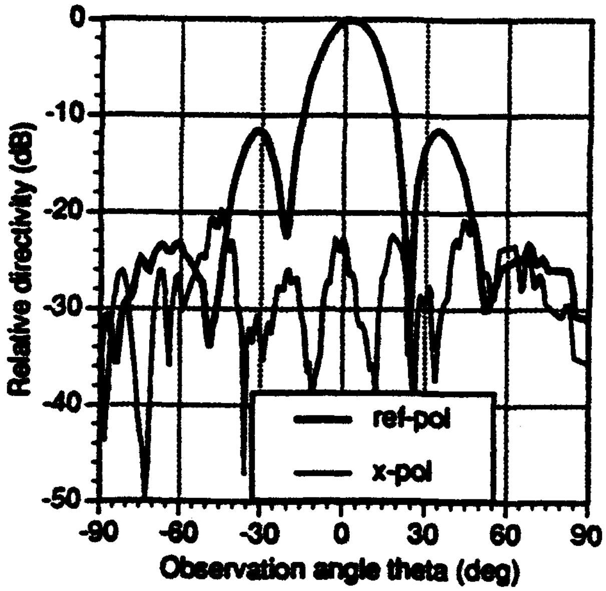


Figure 8. Phi = 90° cut for the 4 x 4 array in the receive band.

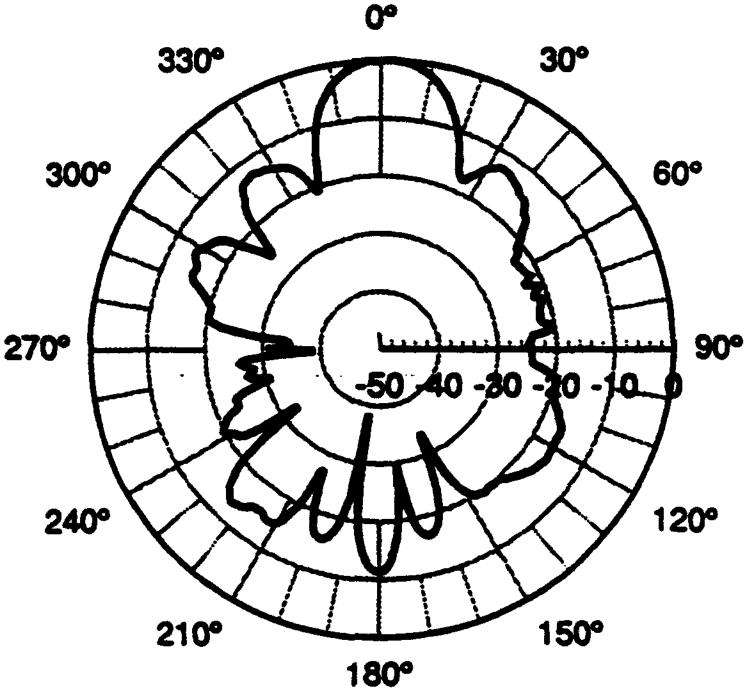


Figure 9. 360° far-field pattern for the 4 x 4 array in the receive band for the phi = 0° cut.

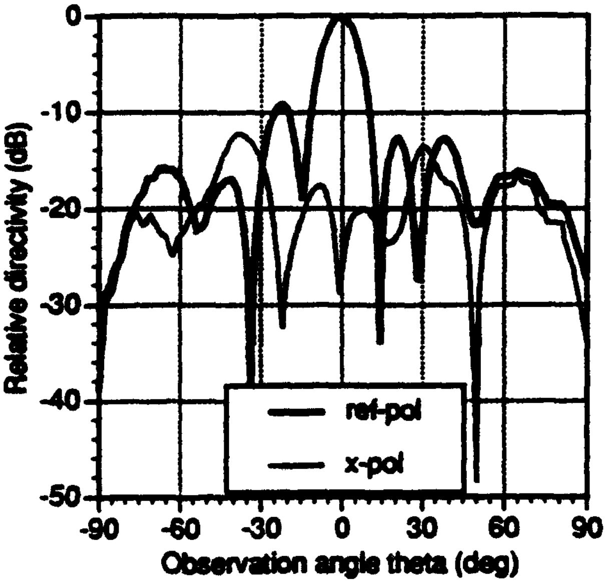


Figure 10. Phi = 0° cut for the 4 x 4 array in the transmit band.

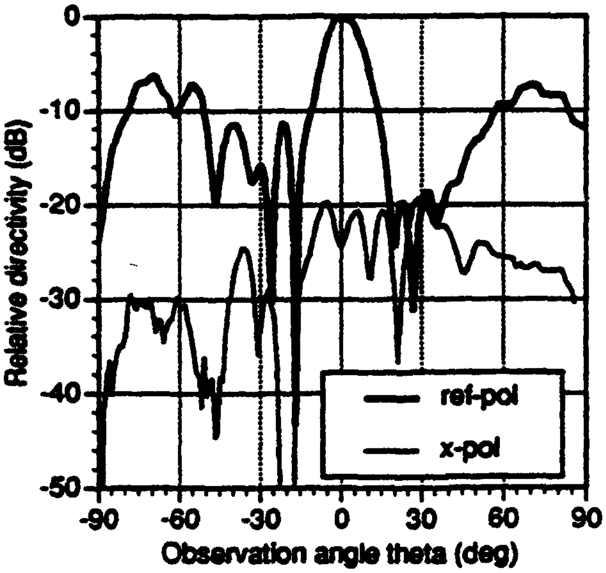


Figure 11. Phi = 90° cut for the 4 x 4 array in the transmit band.

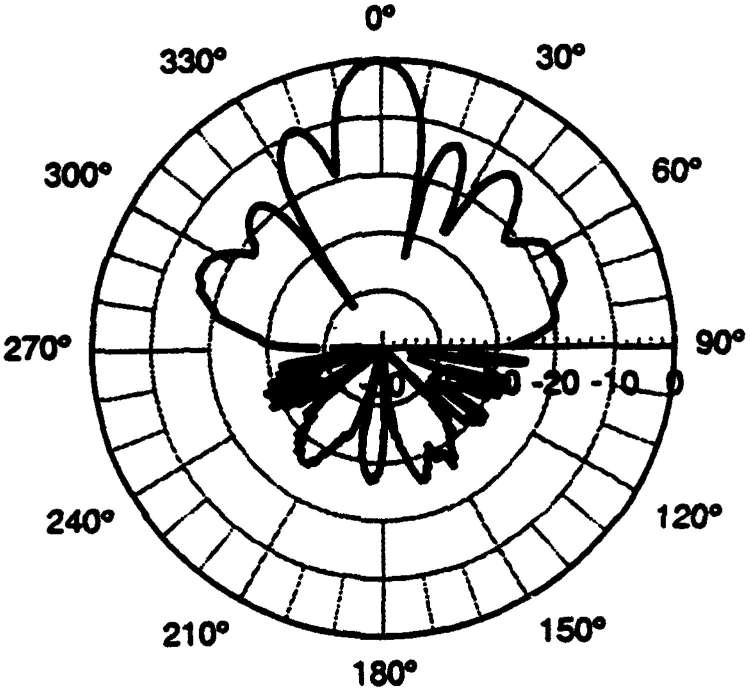


Figure 12. 360° far-field pattern for the 4 x 4 array in the transmit band for the phi = 0° cut.

SCANNING CHARACTERISTICS OF STRIPLINE-FED TAPERED SLOT ANTENNAS ON DIELECTRIC SUBSTRATES

Daniel H. Schaubert
Electrical and Computer Engineering
University of Massachusetts
Amherst, MA 01003

Abstract

Endfire tapered slot antennas are a class of printed antennas that offer the possibility of producing wide-bandwidth, wide-scanning phased arrays that are compatible with integrated circuit fabrication and assembly techniques. Single tapered slot antennas fed by stripline or microstrip can be readily designed to operate over bandwidths exceeding two octaves and arrays have also demonstrated wideband operation in configurations that permit grating-lobe-free scanning over wide angles. However, mutual coupling in the array environment has a dramatic effect on antenna performance and many developers of small to moderate size arrays have observed impedance and pattern anomalies reminiscent of scan blindness. Few examples of these problems have been reported in the literature and few analyses have attempted to assess these effects and to develop design procedures that eliminate or reduce these deleterious effects. This paper will present results from one such numerical study that is attempting to determine the physical phenomena that underlie the most catastrophic of these effects, i.e. scan blindness. The behavior of the antenna arrays is quite good at all but a few isolated frequency bands where a blindness appears and sweeps through a

range of scan angles. Some of the results obtained so far indicate that the grid spacings and the metalization pattern of the antenna are of primary importance and that the dielectric substrate modifies, but may not drastically change, the performance of tapered slot antenna arrays.

1. Introduction

Endfire tapered slot antennas, also known as Vivaldi or notch antennas, have been demonstrated to efficiently radiate wide-bandwidth signals[1-4] and to provide wide-angle scanning capability in wide-band phased arrays[5]. The antennas are unique in their ability to operate efficiently in arrays with element spacing varying from less than 0.2 wavelengths at the lower portion of the operating band to approximately 0.5 wavelengths at the upper portion.

Endfire tapered slot antennas are frequently fabricated by etching a flared slot pattern in the metalisation of a microwave substrate, but self-supporting metallic fins are used also. The narrow slotline at the antenna's feed can be coupled to a variety of transmission lines including microstrip, stripline and coax. The transition at this point is important for achieving wide-band operation.

Single elements demonstrate a wide VSWR bandwidth. Their radiations patterns are very broad when the elements are electrically small and are moderately directive (beamwidths of 30°–60° are common) when the elements are a few wavelengths long with aperture heights of one or two wavelengths. The polarisation of single antennas is purely linear in the principal planes if the antenna structure is symmetric, e.g. a bilaterial slotline antenna fed by stripline. However, relatively high levels of cross-polarisation are observed in the diagonal planes[6-7].

When used in arrays, mutual coupling between tapered slot elements can be quite large, especially when the elements are electrically small and closely spaced as in the lower portion of the operating band of a typical wide-band array. Nonetheless, the antennas can perform quite well, exhibiting VSWR < 2 over wide bandwidths and scan angles with active element patterns that approximate the ideal $\cos\theta$ shape. The polarization of the field radiated from large arrays mimics that of single elements, i.e., linear in the principle planes and high cross-polarization in the diagonal plane[8]. Unfortunately, however, arrays of tapered slot antennas can exhibit scan blindnesses. Because these blindnesses are so catastrophic to array performance, it is important that they be identified early in the design cycle and that corrective action be taken. To accomplish this, the phenomena involved in creating the blindness should be understood and, hopefully, predicted by relatively simple means, as for microstrip arrays[9].

This paper presents several case studies of single-polarized, infinite arrays of tapered slot antennas with and without dielectric substrates. The studies are performed by using method of moments analyses that have been developed specifically to treat these types of antennas[10-11]. The goal of the studies is to understand the fundamental principles that determine the behavior of these antennas, especially scan blindness, and to characterize these principles in ways that allow engineers to successfully design arrays. One class of blindnesses has recently been characterized in this way[12], but other blindnesses remain illusive. However, even these blindnesses behave in a somewhat predictable manner and case studies can be enlightening to designers who must deal with these problems.

2. Case Studies of Scan Blindness

All of the results presented here relate to arrays of single-polarized elements in a rectangular grid, an example of which is depicted in Fig. 1. The shape of the radiating elements may vary from that depicted and, as noted above, the antennas can be fed in a variety of ways. Also, self-supporting structures without dielectric have been studied. In most cases they behave similarly to a printed antenna having the same size, shape, and grid spacing.

2.1. *E-plane Blindness When* $a/\lambda_o > 0.5$

Scan blindness in phased arrays results from mutual coupling effects as is well-known. For example, the active reflection coefficient, Γ_a, of an element in an infinite array is[13]

$$\Gamma_a = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{mn} e^{-jkuma} e^{-jkvnb} \qquad (1)$$

where C_{mn} is the coupling coefficient from the element located at $x = ma$, $y = nb$ to the element located at $x = 0$, $y = 0$, $k = 2\pi/\lambda_o$, $u = \sin\theta_o \cos\phi_o$, $v = \sin\theta_o \sin\phi_o$, and (θ_o, ϕ_o) are the beam pointing directions. The coupling factors are independent of scan angle, but the exponential terms in (1) clearly change as the scan angle changes. In some unfortunate cases, at a particular angle or set of angles, the terms on the right-hand side of (1) add together to produce a reflection coefficient having a magnitude of one and, thus, the array is "blind". However, it is sometimes more enlightening to view the scan blindness as a forced resonance of some guided mode traveling across the aperture of the array. In the work of Pozar and Schaubert[9], this point of view led to a relatively simple way to predict scan blindness in microstrip arrays.

For tapered slot arrays of the type depicted in Fig. 1, one class of blindnesses has been identified with a guided wave on a corrugated surface derived from the array[12]. These blindnesses occur in the E-plane of the array and exist only when the H-plane separation of the adjacent metalized antenna surfaces exceeds $\lambda_o/2$. In Fig. 1, the separation $a - t$ must be greater than $\lambda_o/2$. In that case, an electric field having a y component (the dominant polarization of the array) can exist in the air-filled region between the metalization that forms the antennas. For certain critical scan angles in the E-plane, this field takes a form similar to that of a parallel-plate waveguide that has a short-circuit plane at $z = -d$ and resonants with the $m = -1$ Floquet mode in the exterior region to create a guided mode of the type that can be supported on the corrugated surface in Fig. 2. The propagation constant in the y direction is only slightly affected by the shape of the antenna slot, so the angle of blindness can be estimated from the ideal corrugated surface depicted in Fig. 2, i.e., with thin, unbroken corrugations of depth d and separation $a - t$. An example of the actual angle of blindness and the angle predicted by the simple guided wave model is shown in Fig. 3. Although the agreement is not perfect (indicating some loading effects on the guided mode caused by the antenna slot), the trend is accurately predicted, including a cut-off of the mode and no blindness for $(a - t)/\lambda_o$ less than about 0.55.

Further evidence of the validity of this simple model for the blindness is contained in Fig. 4, where actual and predicted values of the blind angle are compared for a tapered slot antenna. In this case, the prediction is seen to be reasonably accurate over the range of frequencies where the blindness occurs. This structure has no dielectric and is fed by a voltage source across the narrow portion of the slot. Finally, the blindness is seen in Fig. 5 to be relatively insensitive to the thickness and permittivity of the substrate, which further supports the theory that a guided wave in the air-filled region is causing the blindness.

It has been observed that this class of blindnesses exists only in the E-plane of the array. Furthermore, since it only occurs for $a - t > \lambda_o/2$, it is not important for arrays that scan to wide angles in the H-plane and, thus, must have H-plane spacings no greater than $\lambda_o/2$. However, this blindness, which occurs very near to broadside, can be a severe limitation on element spacing for arrays designed to scan only in the E-plane or to scan mainly in the E-plane and cover a limited sector in the H-plane.

2.2. *A Blindness for Grid Spacings Less Than* $\lambda_o/2$

The example presented here illustrates a blindness that occurs for grid spacings *slightly less than* $\lambda_o/2$. The antenna, which is depicted in Fig. 6, is comprised of self-supporting metal fins without a dielectric substrate. At 0.448 GHz, the grid spacing of this array is about $0.48\lambda_o$, an E-plane scan blindness occurs at $\theta = 23°$. This blindness cannot be predicted by the simple corrugated surface model, and it appears to be somewhat different in behavior from that type of blindness. In particular, the array is essentially blind over a portion of the $u - v$ scan space. The active input resistance of the array is plotted in Fig. 7 for the E-plane and two intercardinal planes. The very low values of R_{in} indicate that the array is essentially blind in each of these scan planes. The trajectory of the blind angle in $u - v$ space is shown in Fig. 8. The trajectory follows a path along which $v = \sin\theta_o \sin\phi_o$ is approximately constant, which allows the antenna designer to estimate the usable scan volume after determining only one or two points along the trajectory.

2.3. *Effect of Gaps in Metalization*

When fabricating arrays of antennas integrated with active modules, it may be advantageous to construct units comprised of one radiating element and its associated active components. These units could be individually tested and then plugged into a backplane distribution network to form the array. A ground plane with openings for the radiating elements would complete this array (see Fig. 9), which could be repaired quite easily by replacing defective units. Parfitt and Griffin[14] have recently analyzed an array of this type comprised of printed dipole elements. Although this modular construction allows simpler component testing and repair of defective elements, array performance may deteriorate significantly. The Smith chart plots in Fig. 10 show that the introduction of gaps between the elements causes a blindness to appear at 12.1 GHz. As with the other blindnesses reported in this paper, the metalization is more important than the *dielectric substrate*. Changing the relative permittivity of the substrate from 1 to 3.8 causes the blindness to move only from 14.5 GHz to 10.25 GHz (Table 1).

Table 1. Effect of Dielectric Permittivity on Blindness of CWSA

ϵ_r	Frequency of Blindness (GHz)
1.0	14.50
2.2	12.10
3.8	10.25
6.0	8.70

3. Summary

Although arrays of tapered slot antennas can potentially operate over wide bandwidths and scan to wide angles, the unsuspecting designer can easily encounter severe anomalies in antenna behavior. This paper has presented case studies that illustrate scan blindness. By studying these and other examples, some characteristics of the arrays can be determined and some design guidelines can be inferred.

- One class of scan blindnesses is associated with guided waves on a corrugated surface. These blindnesses occur in the E-plane when the H-plane spacing exceeds $\lambda_o/2$ and they can be predicted by a simple model.
- Other blindnesses occur for grid spacings less than $\lambda_o/2$. These blindnesses depend upon grid spacing and the metalisation pattern of the antenna, but are not strongly dependent on the dielectric substrate.
- Gaps in the metalisation between the antennas can create blindnesses and should be avoided, if possible.

Further work is needed to characterize the performance of taper slot antenna arrays and experiments are needed to verify the characteristics predicted by the numerical studies.

4. Acknowledgement

This work was supported in part by the US Army Research Office under contract DAAL03-92-G-0295.

5. References

1. Lewis, L.R., Fasset, M. and Hunt, J. (1974) A broadband stripline array element, *Digest of 1974 IEEE Symp. Ant. Prop.*:335-337.

2. Gibson, P.J. (1979) The Vivaldi aerial, *Digest of the 9th Eur. Microw. Conf.*, Brighton, UK:120-124.
3. Kollberg, E.L., Johansson, J., Thungren, T., Korzeniowski, T.L. and Yngvesson, K.S. (1983) New results on tapered slot endfire antennas on dielectric substrates, *8th Int'l. Conf. Infrared & mmwaves*, Miami, FL: 12-17.
4. Schaubert, D.H. (1990) Endfire slotline antennas, *JINA '90 Digest*: 253-265.
5. Posgay, J.H. (1992) A wideband sub-array radiator for advanced avionics applications, *1992 Antenna Applications Symposium*, Robert Allerton Park.
6. Schaubert, D.H. (1989) Endfire tapered slot antenna characteristics, *Sixth Int'l. Conference on Antennas and Propagation, ICAP89*, IEE Conference Publication 301, part 1:432-436.
7. Schaubert, D.H. (1989) Radiation characteristics of linearly tapered slot antennas, *1989 IEEE AP-S International Symposium*, San Jose, CA:1324-1327.
8. Schaubert, D.H. (1992) Wide-bandwidth radiation from arrays of endfire tapered slot antennas, *Int'l. Conf. on Ultra-Wideband Short-Pulse Electromagnetics*, Polytechnic Institute of New York.
9. Posar, D.M. and Schaubert, D.H. (1984) Scan blindness in infinite phased arrays of printed dipoles, *IEEE Trans. Ant. Propagat.* AP-32 (no. 6): 602-610.
10. Cooley, M.E., Schaubert, D.H., Buris, N.E. and Urbanik, E.A. (1991) Radiation and scattering analysis of infinite arrays of endfire slot antennas with a ground plane, *IEEE Trans. Ant. Propagat.*, AP-39 (no. 11): 1615-1625.

11. Aas, J.A., Schaubert, D.H., Cooley, M.E. and Buris, N.E. (1993) Moment method analysis of infinite antenna arrays with a ground plane, submitted to *IEEE Trans. Ant. Propagat.*

12. Schaubert, D.H. and Aas, J.A. (1993) An explanation of some E-plane scan blindnesses in single-polarized tapered slot arrays, *Digest of 1993 IEEE Antennas and Propagation Society International Symposium*, Ann Arbor, MI: 1612-1615.

13. Oliner, A.A. and Malech, R.G. (1985) Mutual coupling in infinite scanning arrays, in *Microwave Scanning Antennas*, Vol. 2, R.C. Hansen, Ed., Peninsula Publishing, Los Altos, CA.

14. Parfitt, A.J., Griffin, D.W. and Cole, P.H. (1993) Analysis of infinite arrays of substrate – supported metal strip antennas, *IEEE Trans. Ant. Propagat.*, AP–41 (No. 2): 191-199.

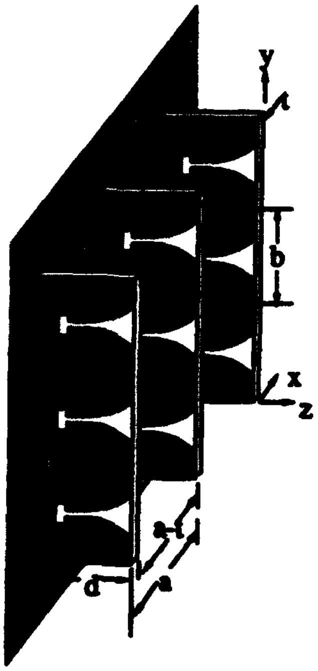


Figure 1. Portion of an infinite array of singly polarized tapered slot antennas.

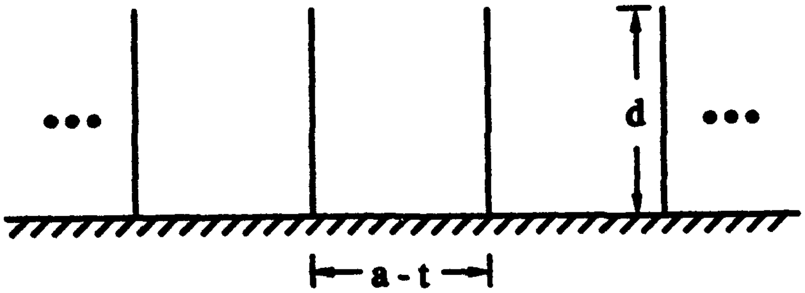


Figure 2. Corrugated surface derived from tapered slot array.

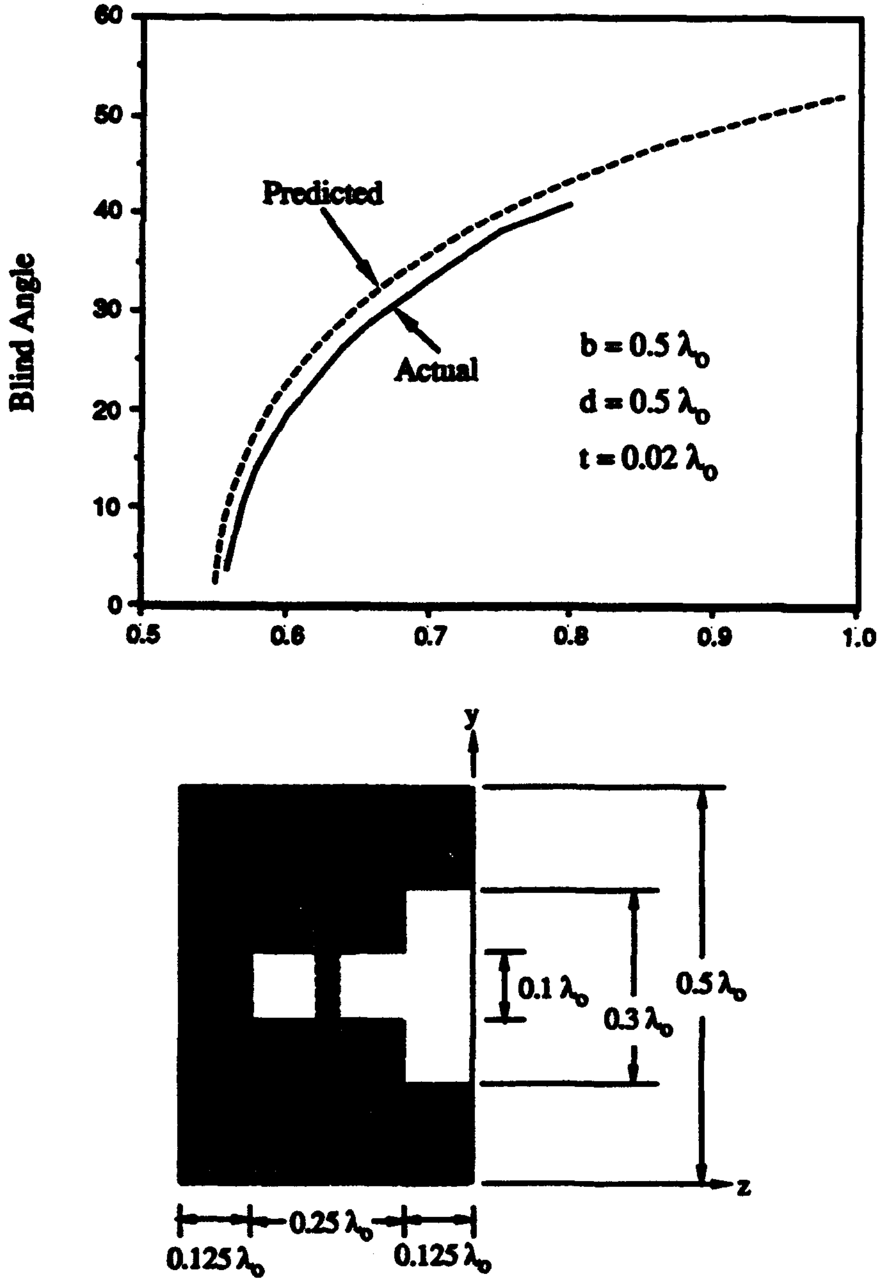


Figure 3. Effect of H-plane separation on angle of blindness for CWSA array with stripline feed. $\epsilon_r = 2.2$.

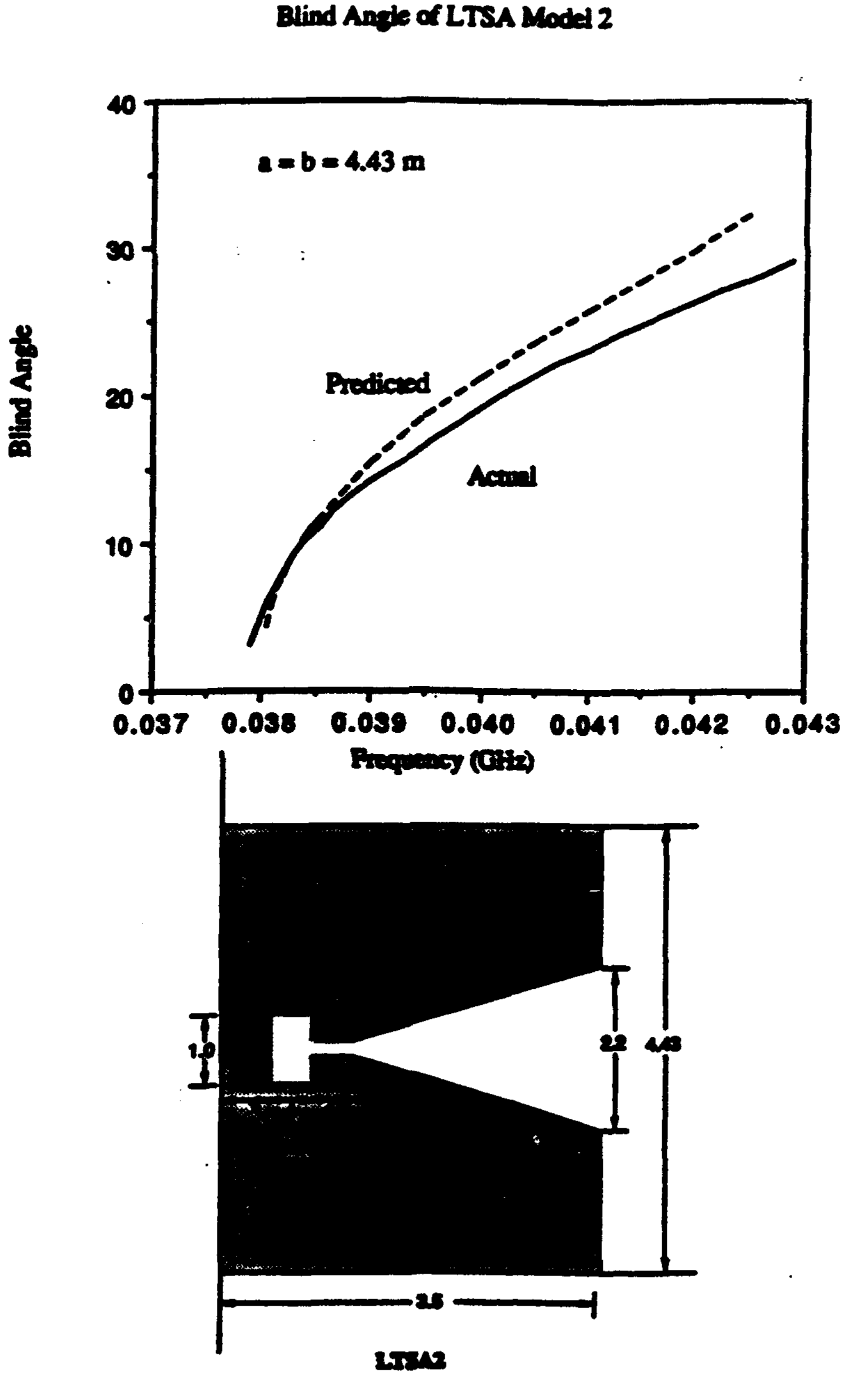


Figure 4. Frequency dependence of blind angle for a tapered slot antenna array with self-supporting fins. $t = 0$, $\epsilon_r = 1.0$. Dimensions are in meters.

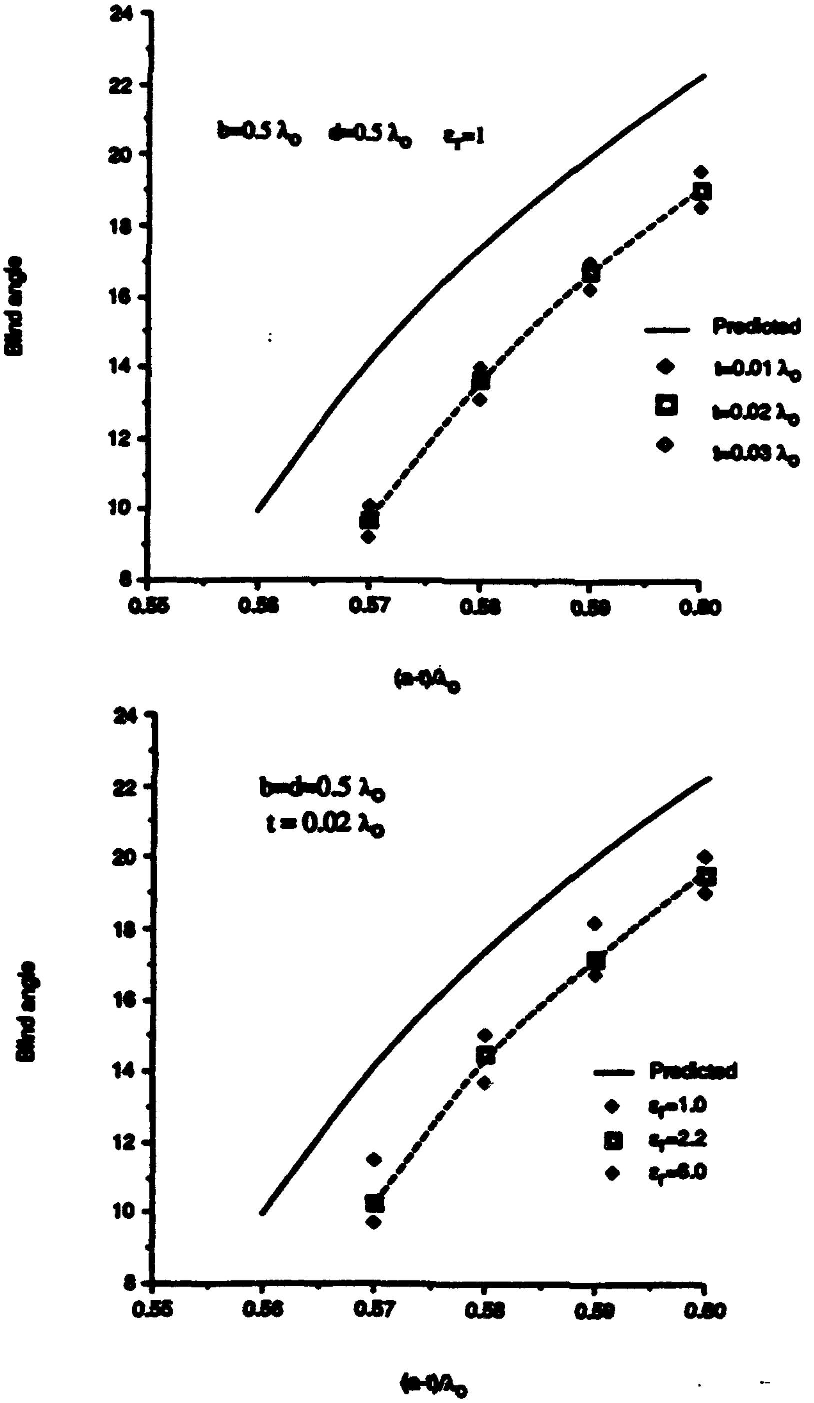
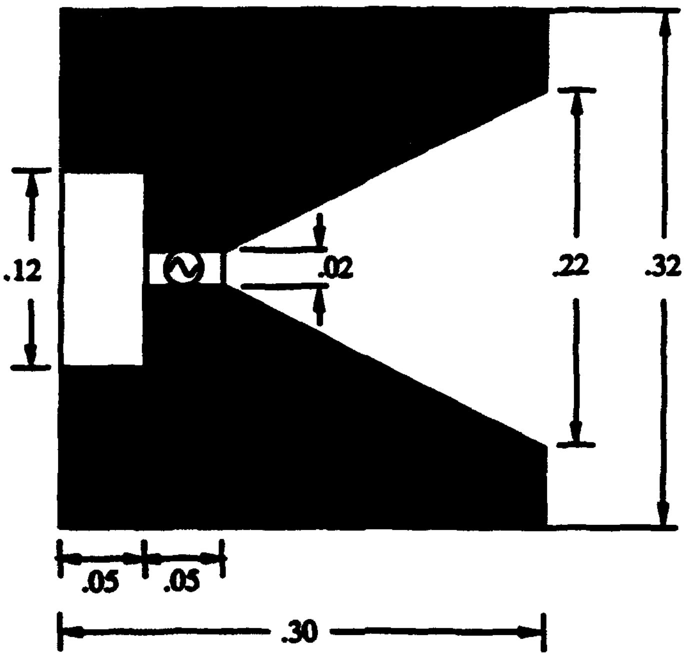


Figure 5. Effect of substrate thickness and permittivity for stripline-fed CWSA shown in Fig. 3.



Antenna 2

Figure 6. Dimensions in meters of linearly tapered slot antenna (LTSA) element.

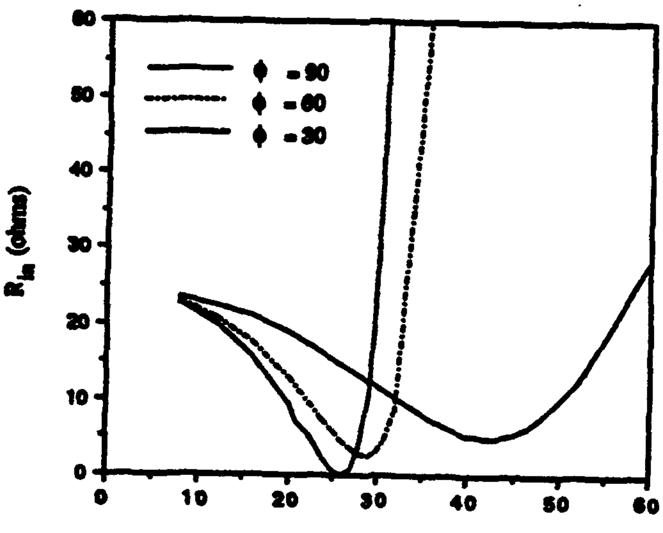


Figure 7. Input resistance in three scan planes of LTSA shown in Fig. 6. Frequency = 0.448 GHz.

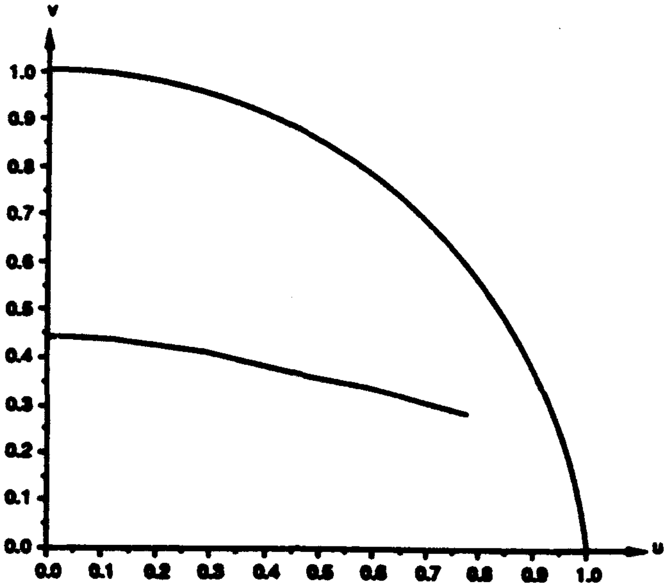


Figure 8. Trajectory in $u - v$ space of input resistance minimum for LTSA shown in Fig. 6. Frequency = 0.448 GHz.

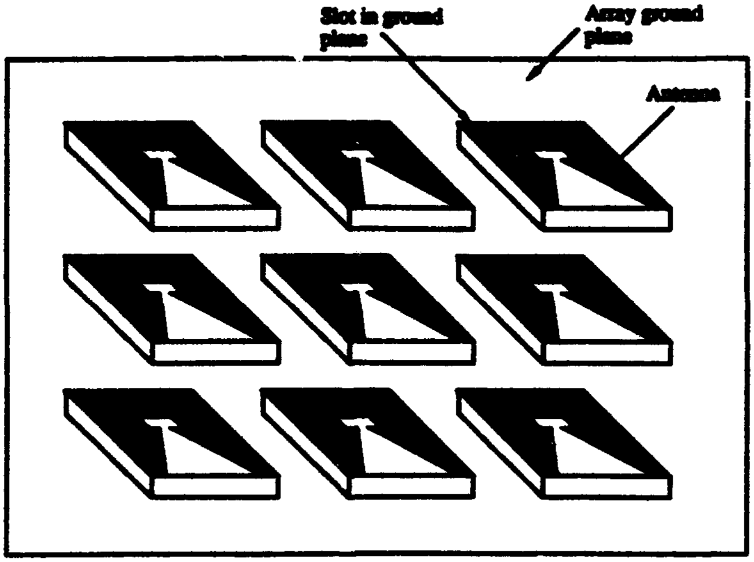

Slot in ground plane
Array ground plane
Antenna

Figure 9. A scheme for building tapered slot antenna arrays by using modules with separate substrates for each antenna.

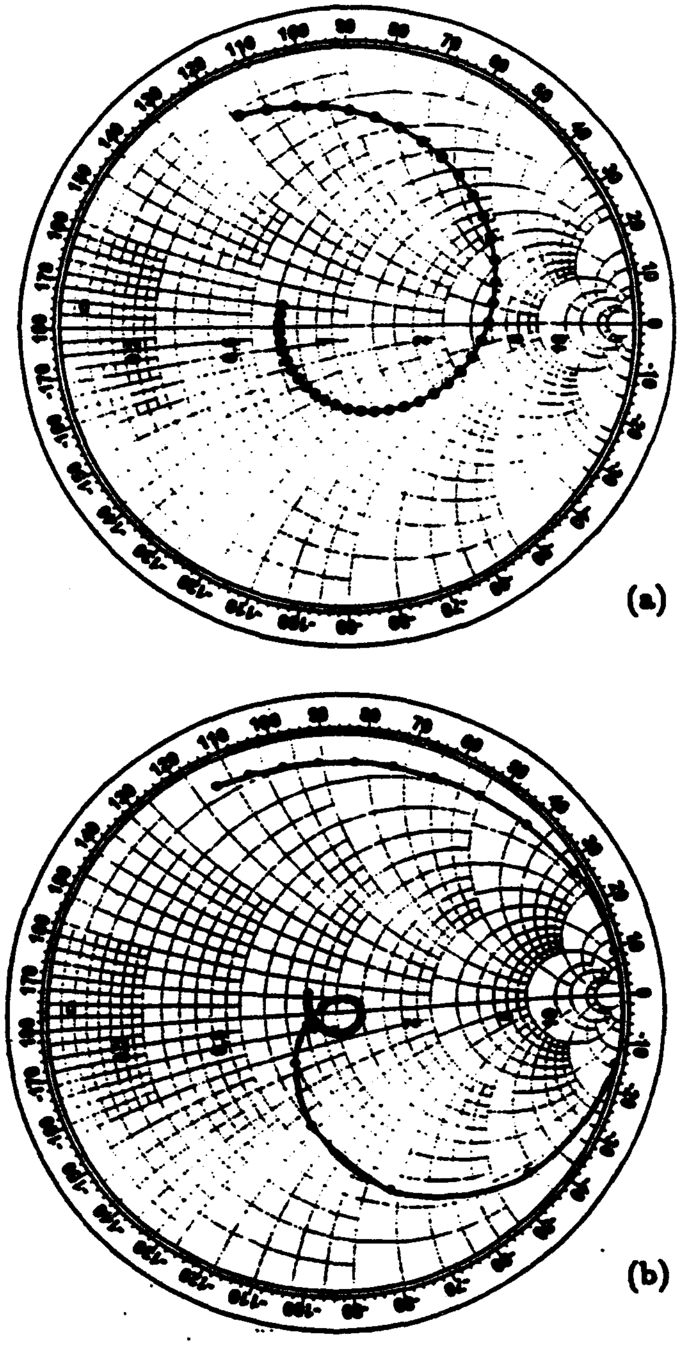


Figure 10. Input impedance of CWSA antenna shown in Fig. 3 at $\theta = 10°$, $\phi = 90°$. $\epsilon_r = 2.2$, $t = 0.5$ mm. (a) Continuous metalization between elements. (b) 0.1-mm-wide gaps separating elements.

A New Waveguide-to-Microstrip Transition

by

Naftali Herscovici

D&M/CHU Technology, Inc.
Whitcomb Avenue
Littleton, MA 01460

August 1993

Abstract Large microstrip arrays exhibit low efficiency mainly because of losses associated with long microstrip transmission lines. One way to overcome this problem is to replace parts of the feeding network with alternative low loss transmission lines, such as the waveguide. The efficiency of the array is also partially determined by the efficiency of the waveguide-to-microstrip transition. In this paper, a new waveguide-to-microstrip transition is proposed. By using the waveguide instead of the microstrip line as a primary feeding network, losses are reduced and also the power handling capability of the array is considerably improved.

1. Introduction

In the past, many different waveguide-to-microstrip configurations were proposed. Some of them are shown in Figure 1. In the transition shown in Figure 1a, a microstrip line is physically connected to the ridge of a single-ridged waveguide, which transforms into a regular waveguide. This transition exhibits a relatively good SWR, however the power transfer efficiency is low and its mechanical characteristics are marginal(the amount of pressure imposed by the ridge on the microstrip line is critical). In Figure 1b, 1c and 1d the coupling is achieved by an aperture in the large/transversal wall of the waveguide. The main drawback of this configuration resides in its low coupling level. This limits the applicability of the transition which can be used only in very large arrays, where the required coupling-per-element is compatible with the power coupling levels reported in [1].

In this paper, a new waveguide-to-microstrip transition is presented. The measured coupling is higher than ones reported in the past, and the its losses are shown to be low. Two microstrip arrays using this transition were built and tested. The measured characteristics of these two antennas prove the improved performance and widen the applicativity of the proposed transition.

2. The waveguide-to-microstrip transition

Similar to the waveguide-to-microstrip transitions described in [1], the coupling of the improved waveguide-to-microstrip transition is achieved through an aperture. However here, the aperture is a rectangular slot in the *narrow* wall of the waveguide and the slot axis is *parallel* to the waveguide axis (Figure 2). This geometry is similar to the classical waveguide slot arrays, only in this case, the slots are used as a means of power transfer rather than as radiating element. A four-port coupler similar to the one described in Figure 2 was built and tested. The waveguide is WR75 type and the microstrip line is a 50Ω line printed on a 31 mil thick dielectric substrate having a relative permittivity of 2.22 (Duroid 5880). Two cases were tested:

1. Port II terminated in a matched load (S_{31} is shown in Figure 3).
2. Port II terminated in a short (S_{31} is shown in Figure 4).

The first test measured the coupling plus the losses of the transition, and the second test to measured the transition efficiency. The coupling level is quite constant across the band (here 10 to 15 GHz) and does not exhibit any resonant type of behavior as shown in [2]. For this particular set of parameters, the waveguide-to-microstrip coupling is found to be about 10 dB. This level of coupling is higher than the one reported previously in the literature and is encouraging since it allows the use of waveguide as feeding networks even in small microstrip arrays where power handling is of concern.

Assuming that the dielectric losses and conductive losses are very small, the main source of losses is the slot radiation. The loss of this transition was computed from the S-parameter measurements of the second circuit and found to be about 0.5 dB.

$$\text{Loss} = 1 - |S_{11}|^2 - |S_{31}|^2 - |S_{41}|^2 \qquad (1)$$

3. Applications

A. The 2×2 microstrip array

Two 2×2 microstrip arrays similar to the one shown in Figure 5 were built and tested: one fed by a microstrip line and one fed by a waveguide terminated in a short. The H-plane patterns of the two arrays are shown in Figure 6. The patches are 2.2×2.2 cm and are printed on a 31 mil thick dielectric substrate having relative permittivity of 2.22 (Duroid 5880). The distance between the two patches (in the horizontal plane as well as in the vertical plane) is $0.5\times\lambda_O$ The waveguide is WR187 and in both cases the slot in the array ground-plane is 1 cm×0.1 cm. The microstrip fed array is well matched at 4.46 GHz and the waveguide fed array at 4.38 GHz. This shift in frequency is due to the different contributions of the feeding transmission lines to the reactance of the transition self impedance. As in other slot transitions, this one has the equivalent circuit shown in Figure 7, where

Y_C - is the coupled microstrip self admittance.
Y_{ac} - is the aperture (slot) self admittance looking in the microstrip direction.
Y_{aw} - is the aperture self admittance looking in the waveguide direction.

Assuming the same distribution for the electric field in the aperture (non-resonant slot), the behavior of Y_C and Y_{ac} in this case is identical to the one reported in [2]. The difference between Y_{aw} and Y_{af} in [2] seems to be the source of the shift in frequency mentioned above. These two quantities are given by:

$$Y_{af} = \int_{S_{slot}} \int_{S_{slot}} E_x^{slot}(y,z) G_{xy}^{EMM}(y,z,y_o,z_o) E_x^{slot}(y_o,z_o) ds\, ds_o \qquad (2)$$

$$Y_{aw} = \int_{S_{slot}} \int_{S_{slot}} E_x^{slot}(y,z) G_{xy}^{EMW}(y,z,y_o,z_o) E_x^{slot}(y_o,z_o) ds\, ds_o \qquad (3)$$

where $\tilde{G}^{EM}(y,z;y_O,z_O)$ (*M*-microstrip line, *W*-waveguide) are the Green's functions associated with the feeding transmission lines.

The backlobe appearing in the waveguide fed array is due to diffraction from the waveguide flanges which were too close to the patch. Unlike in the microstrip fed array where the slot is the source of the back radiation, in the waveguide fed array there is no direct source for back radiation.

B. The 2×4 microstrip array

In this case, two 2×2 subarrays of the type described above are fed in series by a WR187 waveguide. To obtain a symmetric excitation of the two subarrays, the length of the slot closer to the waveguide excitation is about 10% smaller than the second one. The distance between the centers of the two slots is λ_g, which is about 9.5 cm. This distance is three times $0.5\times\lambda_0$ as it should be, to insure equal spacing for all the radiating elements. The H-plane pattern is shown in Figure 8.

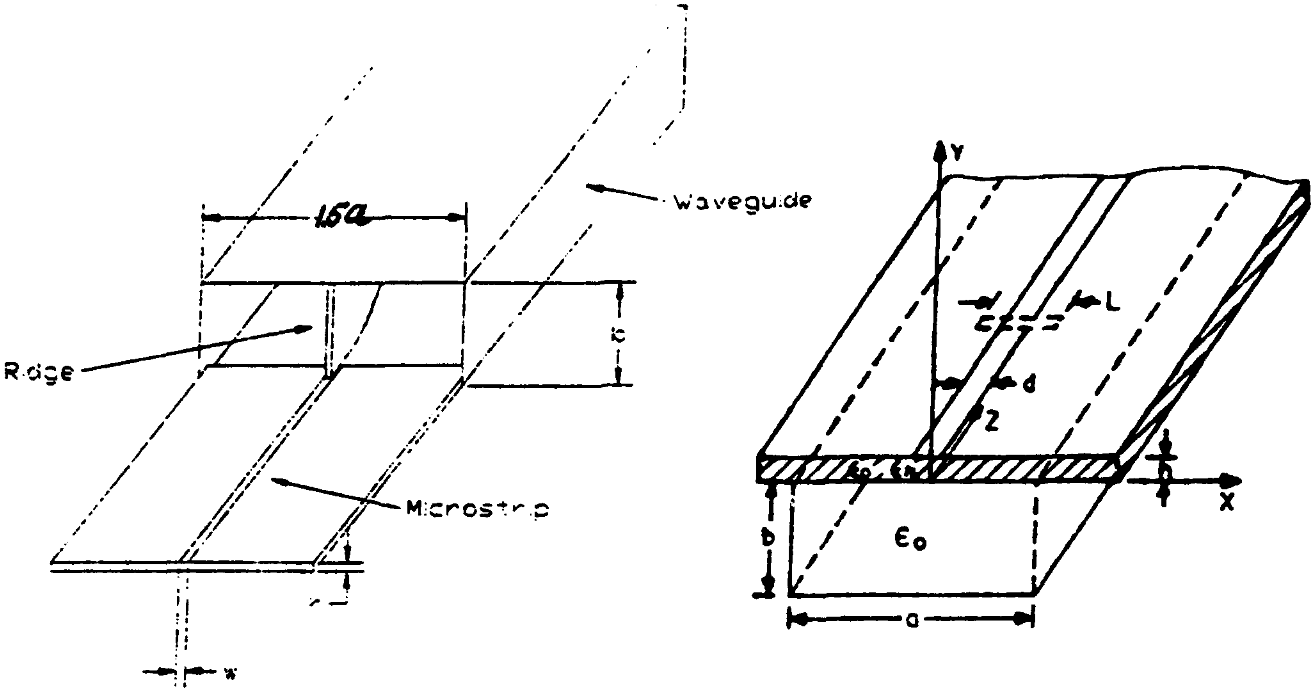
4. Conclusion and future work

A new waveguide-to-microstrip transition was presented. The coupling levels are considerable higher than ones previously reported in the literature. This transition can be used in the integration of primary feeding network with microstrip sub-arrays increasing the array efficiency and power handling capabilities. Two microstrip arrays were built and tested. The modeling of this transition, which is currently in work, will allow a detailed design of such an antenna.

References

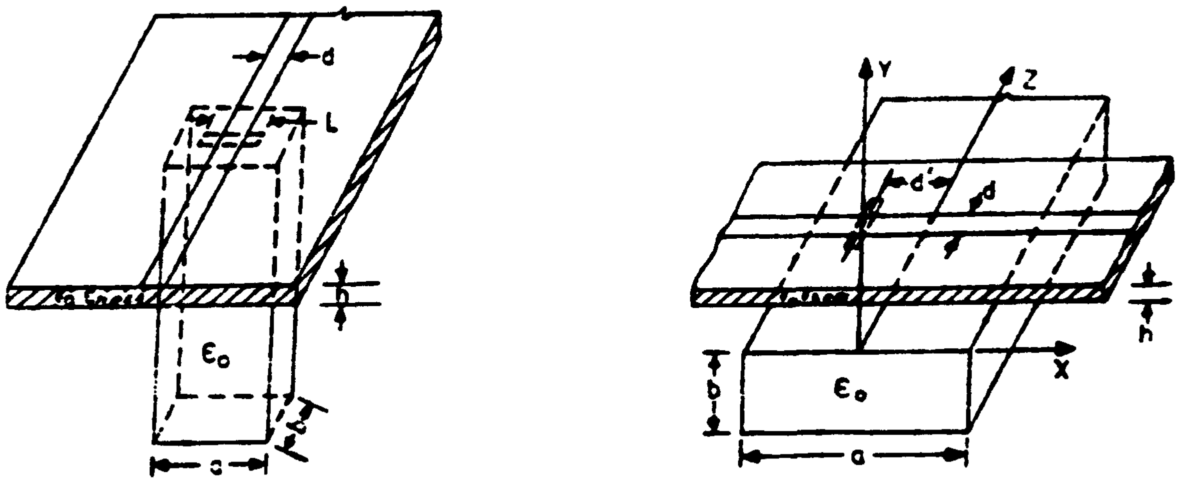
[1] J.S.Rao, K.K.Joshi and B.N.Das, *"Analysis of Small Aperture Coupling Between Rectangular Waveguide and Microstrip Line"*, IEEE Transactions on Microwave Theory and Techniques, Vol.29, No.2, February 1981, pp.150-154.

[2] Naftali Herscovici and D.M.Pozar, *Full-Wave Analysis of Aperture Coupled Microstrip Lines"*, IEEE Transactions on Microwave Theory and Techniques, Vol.39, No.7, July 1991, pp.1108-1114.


1.5a
Waveguide
Ridge
b
Microstrip
h
w
Y
L
d
Z
h
X
b
ε₀
a

a. The ridged waveguide transition

b. Slot in the large wall of the waveguide; the microstrip and waveguide axes are parallel [1]


d
L
h
ε₀
a
Y
Z
d'
d
h
X
b
ε₀
a

c. Slot in the transversal wall of the waveguide [1]

d. Slot in the large wall of the waveguide; the microstrip and waveguide axes are perpendicular [1]

Figure 1 Waveguide-to-microstrip transitions.

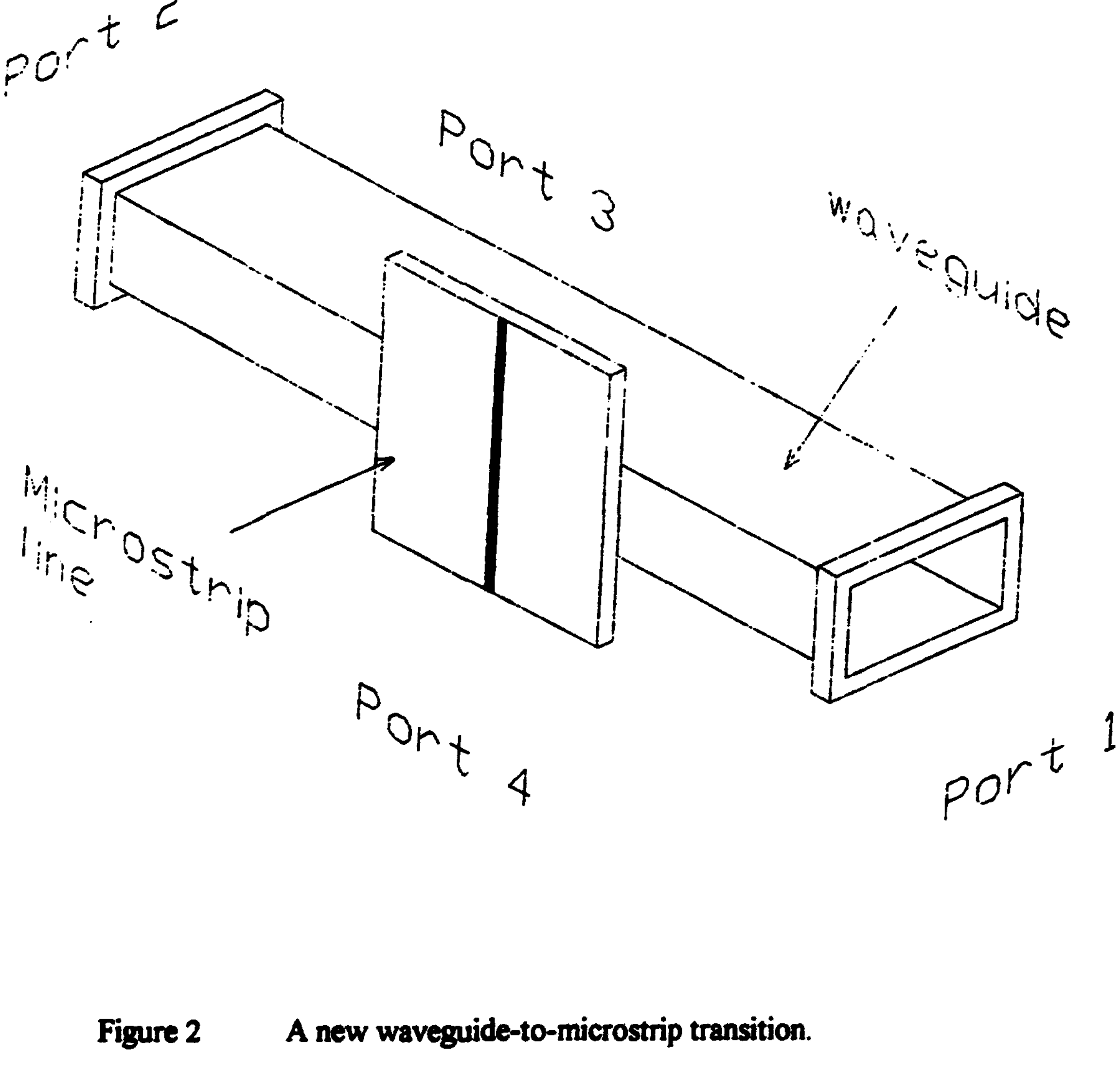

Port 2
Port 3
waveguide
Microstrip line
Port 4
Port 1

Figure 2 A new waveguide-to-microstrip transition.

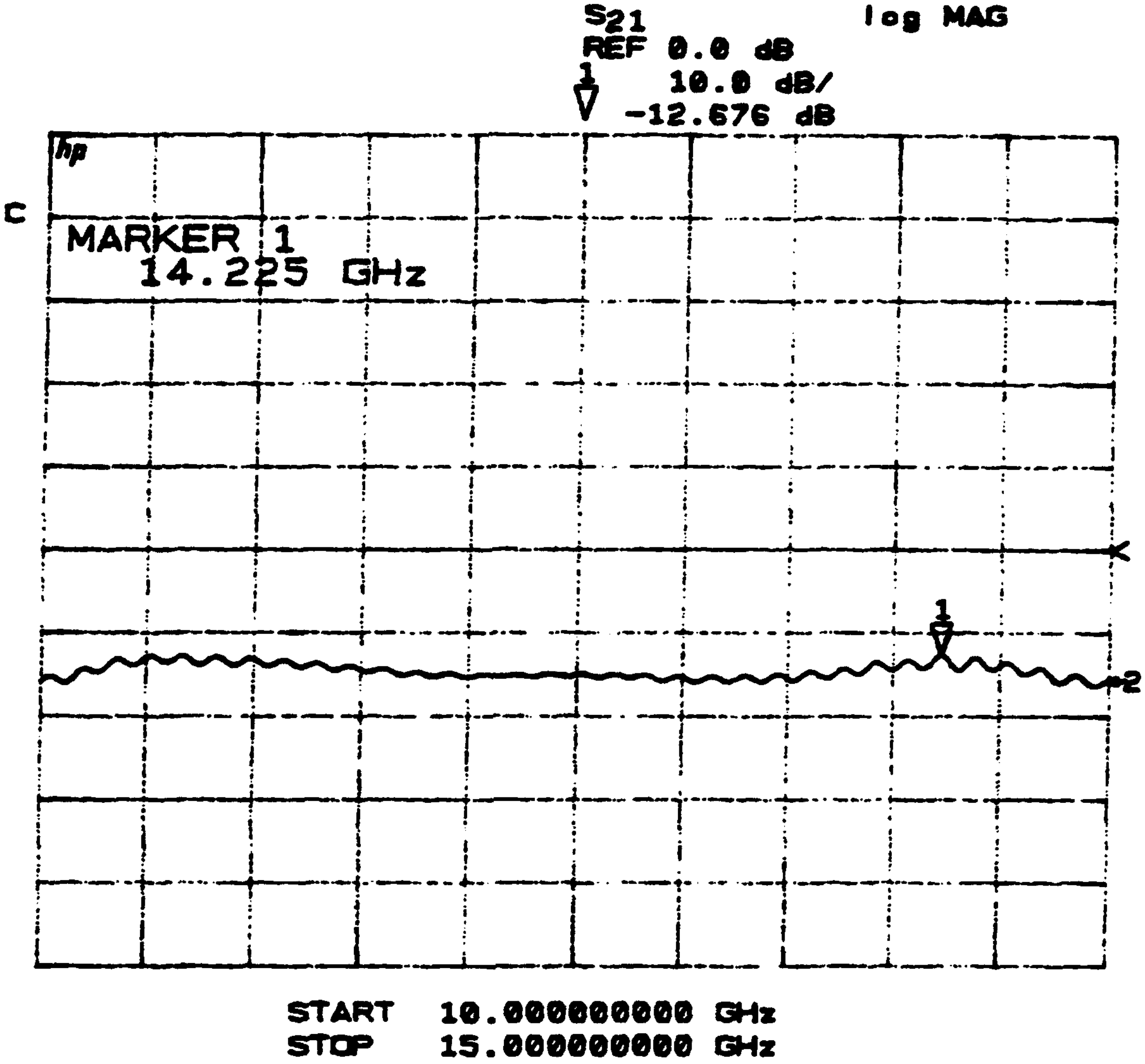


Figure 3 $S_{13}(=S_{14})$ when Port II is matched.

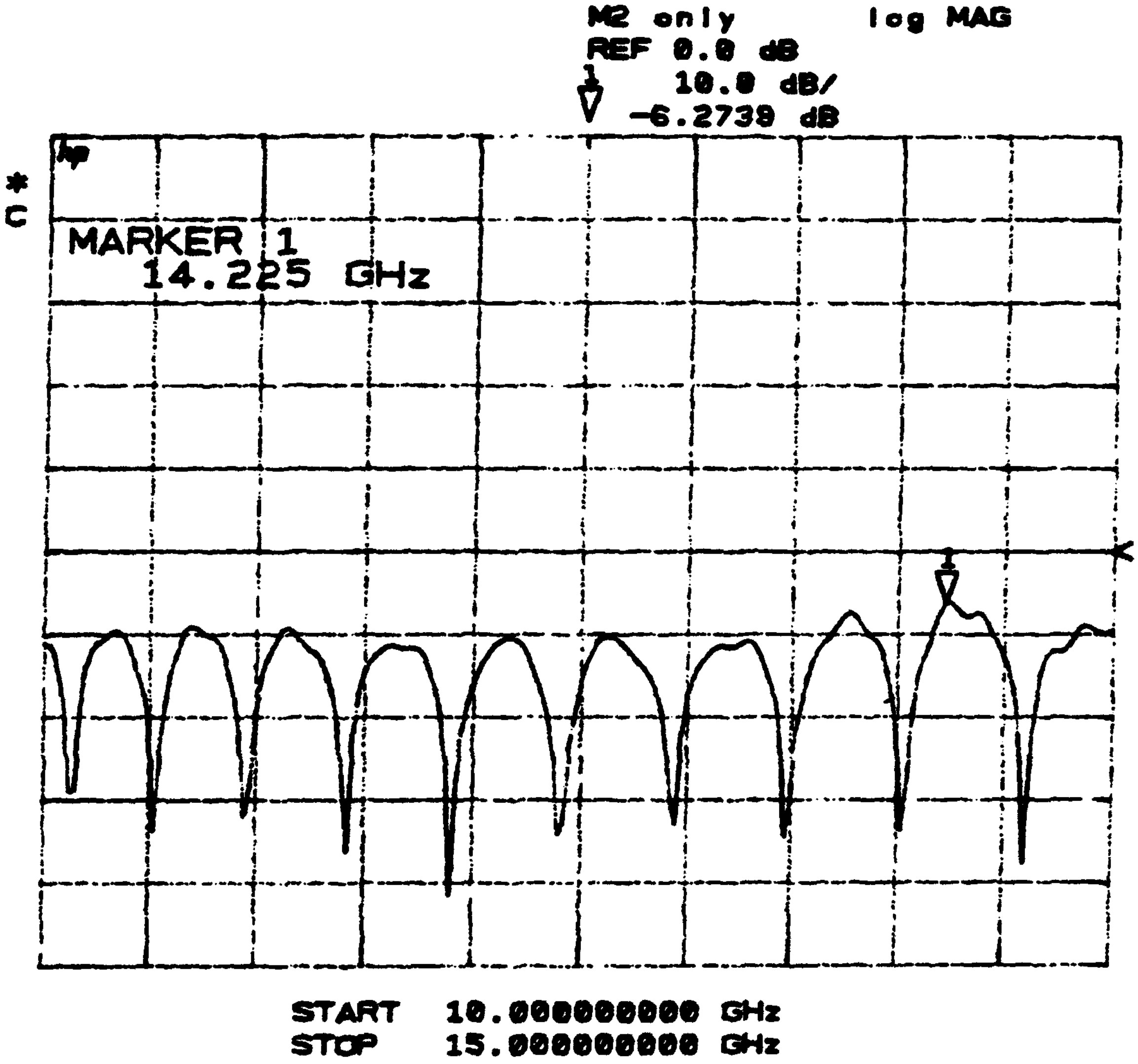


Figure 4 $S_{13}(=S_{14})$ when Port II is terminated in a short.

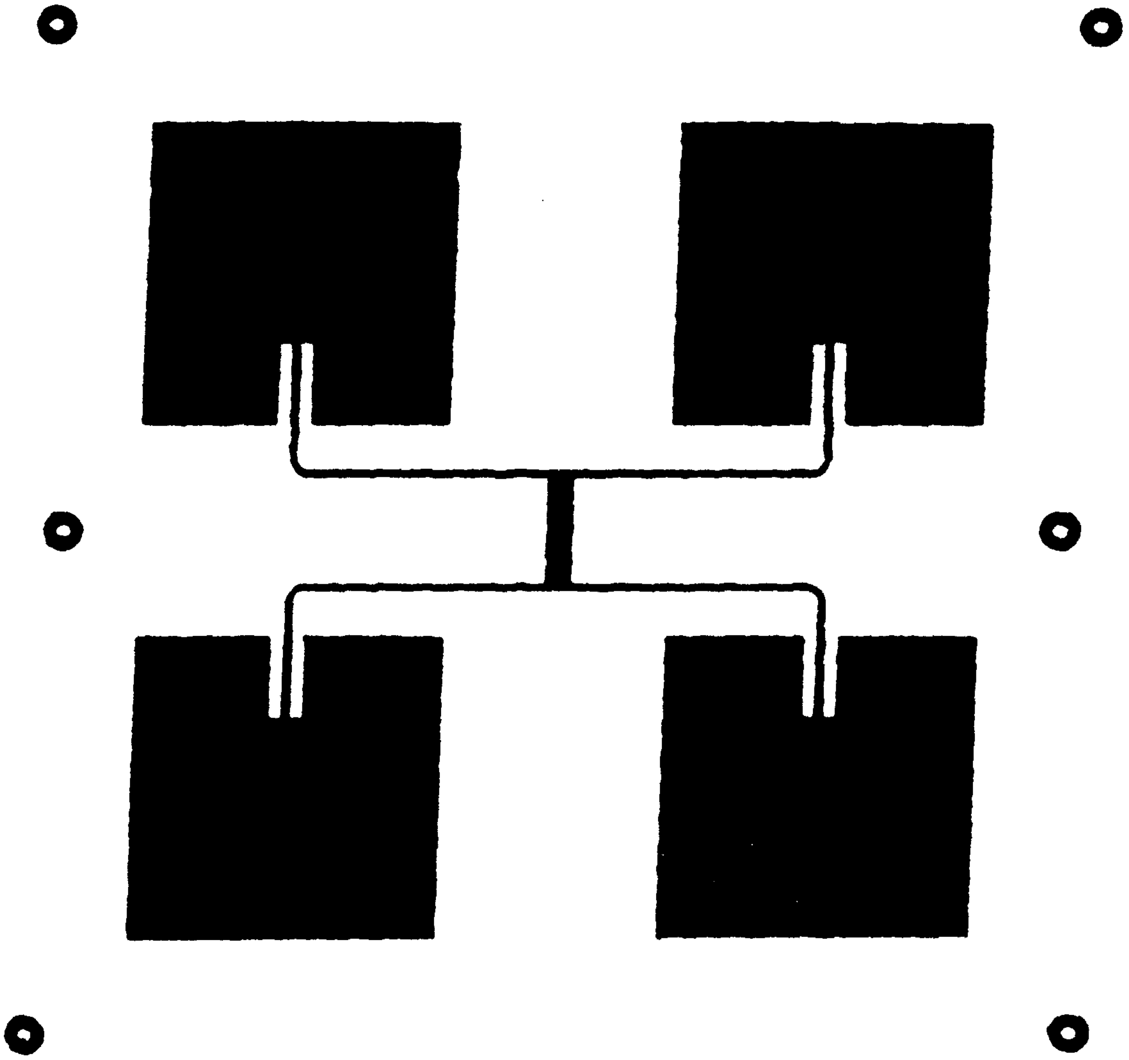


Figure 5 The 2×2 microstrip array - the radiating elements layout.

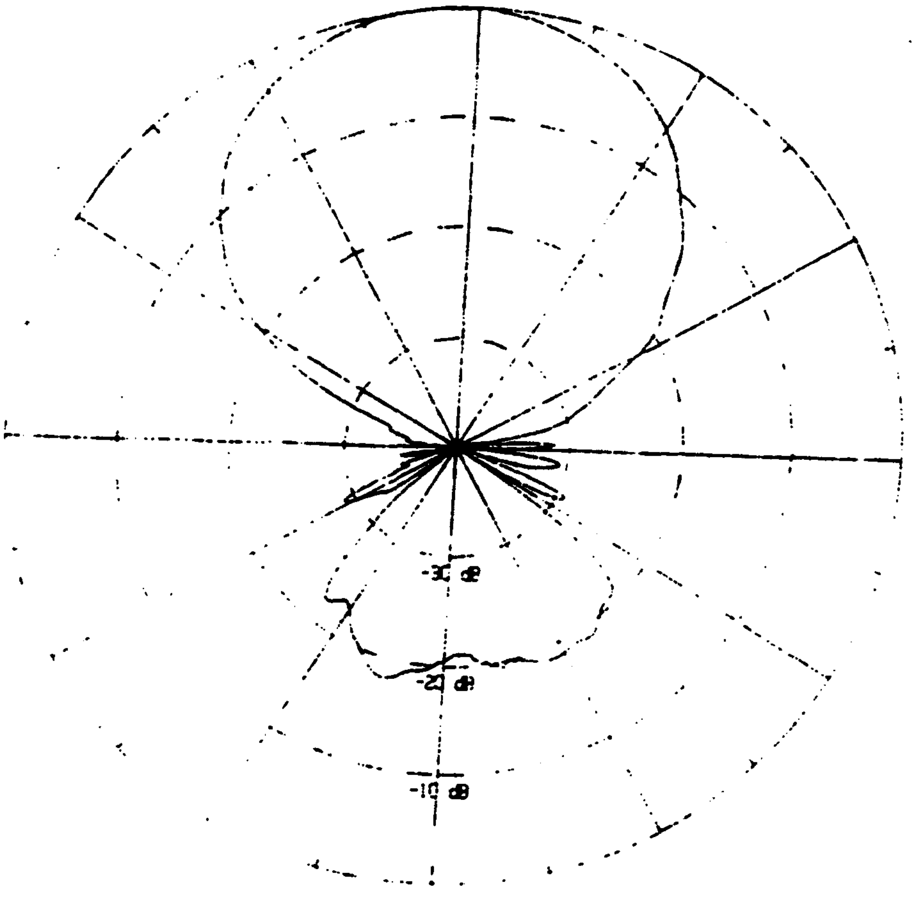


Figure 6a The H-plane patterns of the 2×2 microstrip array fed by a microstrip line.

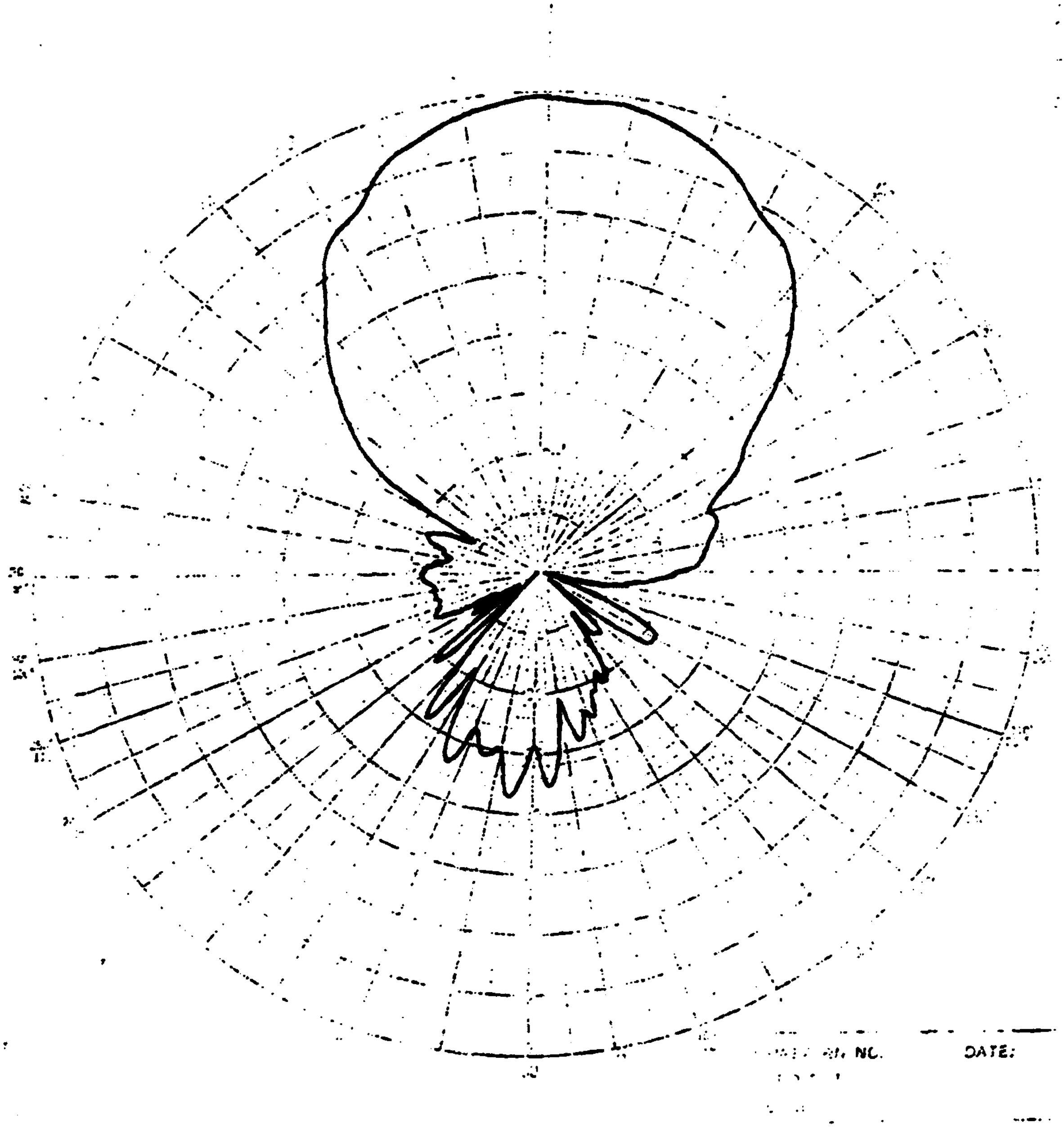


Figure 6b The H-plane patterns of the 2×2 microstrip array fed by a waveguide.

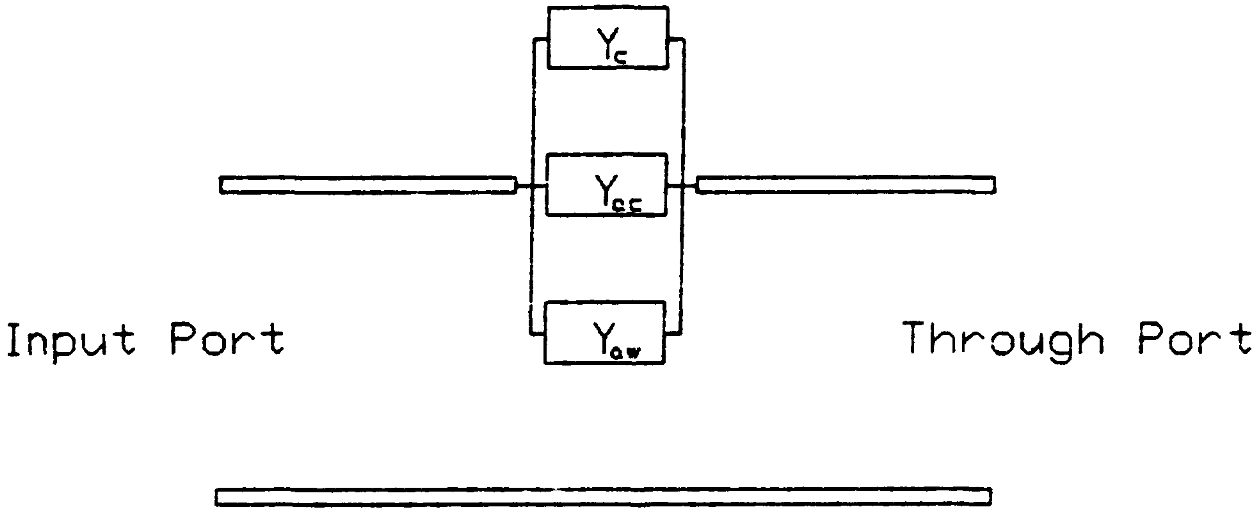

Y_c
Y_{ac}
Input Port
Y_{aw}
Through Port

Figure 7 The transition equivalent circuit.

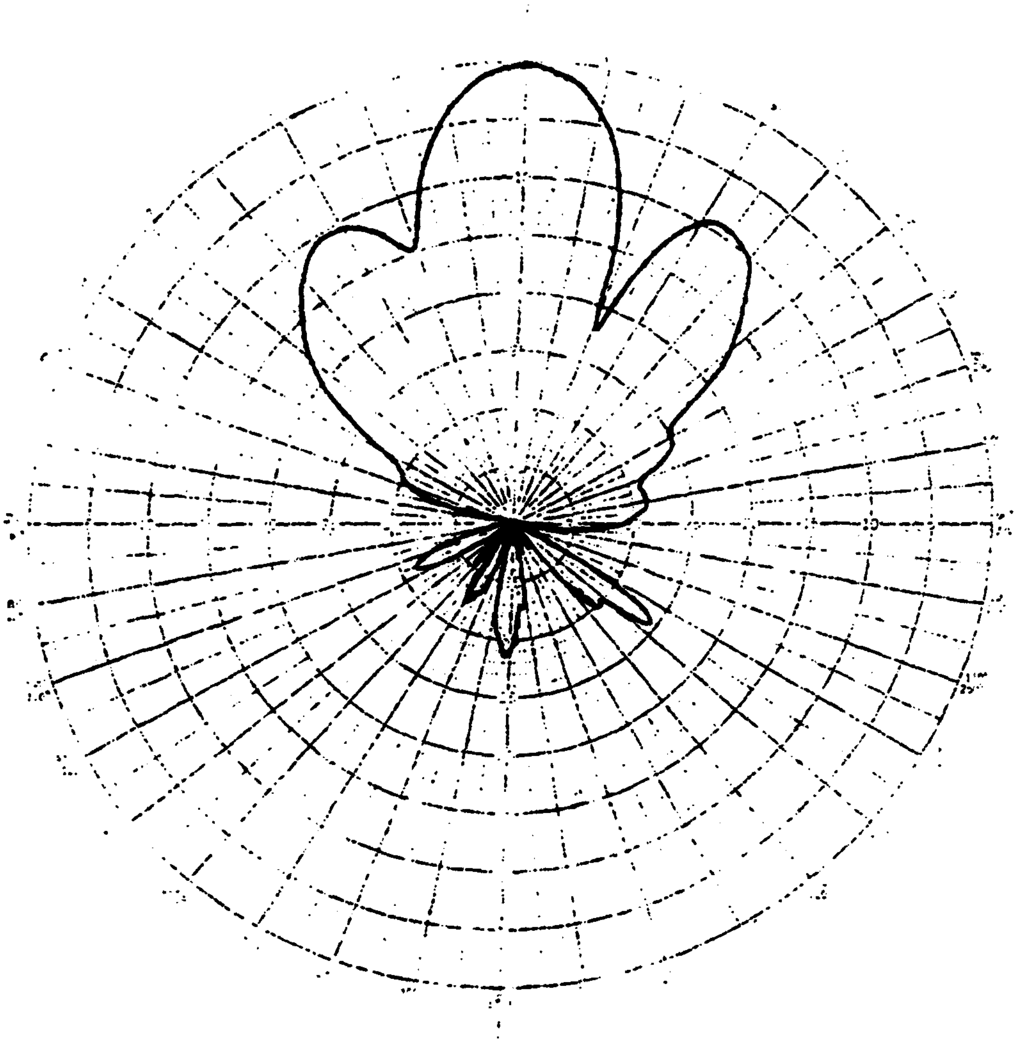


Figure 8 The H-plane pattern of a 2×4 waveguide fed sub-array.

Advances in Neural Beamforming

Major (Dr.) Jeffrey Simmers and Dr. Hugh Southall
USAF Rome Laboratory
Electromagnetics and Reliability Directorate
Hanscom AFB, MA 01731

Terry O'Donnell
ARCON Corporation
260 Bear Hill Road
Waltham, MA 02154

Abstract

The goal of neural beamforming is to develop neural network processing architectures which can perform aspects of digital beamforming with imperfectly manufactured or degraded phased-array antennas. Neural beamforming techniques offer the potential of greatly decreasing manufacturing and maintenance costs of phased-array antenna systems while increasing mission time and antenna performance between repair actions. At last year's symposium, we introduced a neural beamforming architecture, NBF, based on adaptive radial basis functions, which demonstrated the ability to "learn" the single source detection function of an 8-element array having multiple unknown failures and degradations. At that time we had not thoroughly analyzed the single source problem space to insure that our gradient descent approach for converging network weights was robust at finding a near-global minima in the function space. In our subsequent research, we have developed an alternate technique to exactly calculate these weights, using an LMS (least mean square) based pseudo inverse. We incorporated this technique into a new network, LINNET, and compared it's overall performance to the original NBF network on ten sets of X-band antenna measurements under various conditions. We present an analysis of the weight-error surface which reinforces both our hypothesis and our empirical results that the NBF network requires very few iterations to converge to an acceptable solution. We also present an empirical analysis of a critical network parameter, σ, and a preliminary technique which shows promise for predicting the optimal value of this parameter in the future.

1 Introduction

Standard antenna beamforming algorithms, such as monopulse, require highly calibrated antennas because they depend on antenna elements having near-identical behavior and performance. These algorithms typically do not adapt well to uncalibrated antennas or antennas with system or element degradations. As phased-array antennas become larger and more highly integrated into physical structures, the production and maintenance costs to achieve this uniformity become increasingly prohibitive to the use of large antennas for many military and civilian applications.

The requirement for identical elements results from a lack of adaptive beamforming algorithms capable of managing the complexities introduced by non-identical elements with unknown behaviors. Traditional techniques synthesize aperture behavior as a mathematical combination of well-behaved individual element and receiver channel behaviors. Neural beamforming, in contrast, attempts to approximate overall aperture behavior from a finite number of discrete observations of that behavior under varying circumstances. If we assume that antenna aperture behavior is a continuous function, it is theoretically possible to model it with a neural network trained at discrete samples of points along the function. This model can then be used to predict antenna behavior at other points.

We previously presented a neural beamformer, NBF, based on an adaptive radial basis function network, which was successful at approximating the single source location function with measurements from a degraded eight-element linear X-band antenna [6]. Our experimentation with that network led us to question whether the gradient descent function used to adjust internal network parameters was, in fact, achieving a global minima in the output error space or simply falling into local minima.

In this paper, we present a new variation on the NBF network, called LINNET, which uses a linear algebra technique to exactly solve for one set of internal network parameters, instead of an iterative gradient descent technique. LINNET uses an LMS-based pseudo inverse to solve exactly for the output weights to minimize the network error for any particular training set. By developing LINNET and comparing it's performance to NBF across a variety of antenna measurements, we were able to qualitatively measure NBF's ability to converge these output weights to an approximate global solution. We present a comparison of LINNET and NBF when the networks are used to locate single sources at 1° angular intervals after being trained with samples at 10° intervals. The networks are compared for ten measurement sets taken under various signal-to-noise and interference conditions. Our results indicate that NBF's gradient descent performance is very close to LINNET's exact solution.

We present an exploration of the weight-error surface and the topography of a typical weight-error map. The shape of this surface confirms our hypothesis and our empirical results that the NBF network requires only a few iterations to converge to an acceptable solution. We also address two different techniques to determine the optimal σ values of the Gaussian radial basis functions.

2 The Adaptive Radial Basis Function Network

This section presents a brief overview of the Gaussian radial basis function neural network architecture used with the eight element linear phased array digital beamforming antenna for adaptive direction finding. More detail regarding the network and the experiment can be found in a previous paper [6].

This experiment is part of a larger effort to achieve real-time processing of received antenna signals independent of antenna dimension or antenna degradation. Other research sponsored by Rome Laboratory indicates that this goal is supported by incorporating a parallel computer into phase-array antenna systems to process antenna element information in parallel. Each node of the parallel computer would receive signal information from one or more antenna elements, process this local data, and send the output to a global processing system.

The requirement for real-time processing independent of antenna dimension generated two specific guidelines for the neural network architecture. First, it is desirable that the network have a processing delay which can be kept constant regardless of the dimension of the inputs. Second, the network should have a minimum number of layers (layers defined by computations that must be performed sequentially). Thus as the number of inputs increases (as the dimensions of the antenna increase), the size of the network layers should grow at the same rate, and no additional layers should be required.

The radial basis function network architecture reportedly satisfies all of these constraints. The mathematical basis for these networks stems from the fitting (approximation) and regression capabilities of radial basis functions [8, 10], combined with a feedforward neural network architecture whose single "hidden" layer should not grow faster than the number of inputs [1, 7]. We initially chose a three layer network architecture, consisting of an input layer, a hidden layer of Gaussian radial basis functions, and an output layer whose nodes compute a weighted sum of the outputs of the hidden layer nodes. A diagram of this network architecture is shown in Figure 1.

As documented in [9], we found that using raw antenna measurements (consisting of either inphase and quadrature signal components or the amplitudes and phases of the signal) as inputs to the network yielded unacceptable network performance. We developed a preprocessing method that determines the phase-difference between radiation received by consecutive array elements and then uses the sine and cosine of this phase difference as inputs to the network. The evolution of this input preprocessing technique is discussed in [9]. With respect to Figure 1, the input nodes broadcast this preprocessed data to the hidden layer nodes.

The hidden layer performs a *projection* from the input space X into a space denoted by Ψ, such that $(\psi_1, \psi_2, ..., \psi_q) \in \Psi$, where q is the number of nodes in this hidden layer. Each hidden node performs a mapping $[-1,1]^n \rightarrow \mathbf{R}$ given by

$$\psi_i = e^{-\sum_{j=1}^{n} \frac{(x_j - m_{ij})^2}{2\sigma_i^2}} \tag{1}$$

where n is the number of components in each input vector, $(x_1, x_2, .., x_n) \in X$ and ψ_i denotes the output of the ith hidden layer node. Based on the analogy that each ψ_i represents a

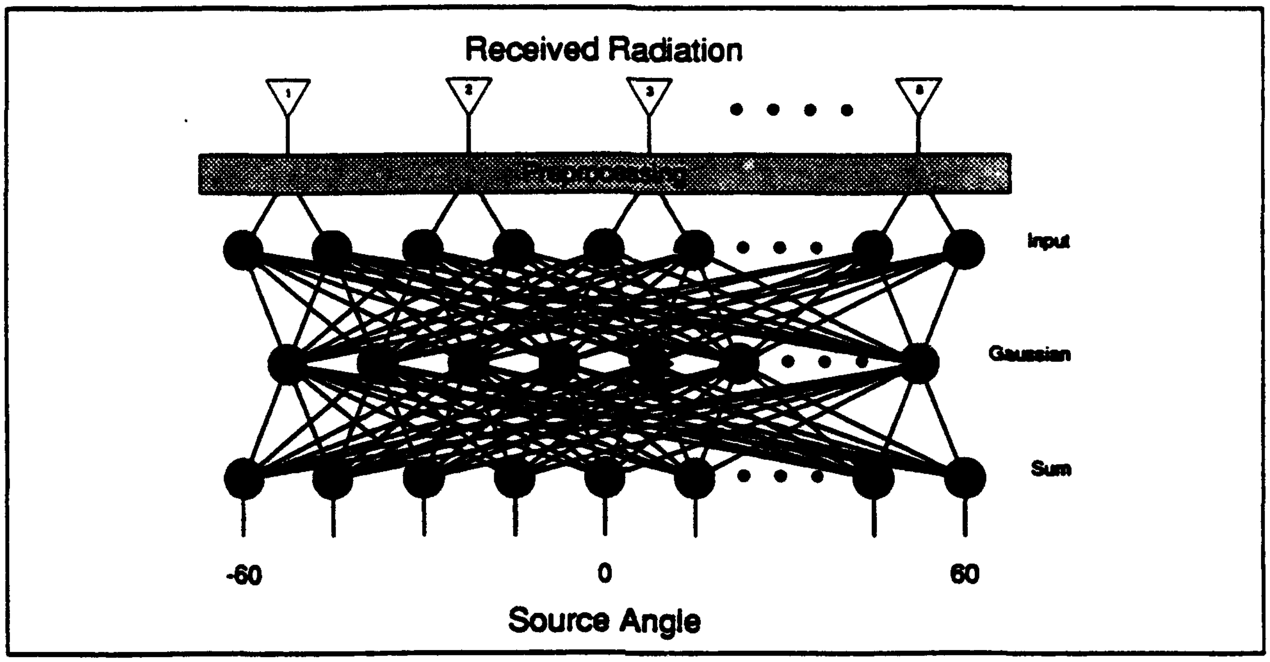


Figure 1: *Adaptive Radial Basis Neural Network architecture, as used in the Neural Beamformer.*

Gaussian function, the σ_i^2 can be considered as the diagonal entries of the covariance matrix that control the spread of each Gaussian "bump". Alternatively, a linear space analogy would be that these σ_i are weight parameters used in computing a weighted norm. That is,

$$\psi_i = e^{-\|x-m_i\|^2_{\sigma_i}} \tag{2}$$

where x and m_i denote the vectors whose components appear in Equation 1, and the σ_i subscript indicates a weighted Euclidean (L_2) norm.

A radial function is one whose evaluation depends upon a radial distance from a given point called the *center* of the function. For each ψ_i, m_i is the vector denoting this center, with components given by $(m_{i1}, m_{i2}, ..., m_{in})$. Thus $\|x - m_i\|^2_{\sigma_i}$ denotes the (weighted) radial distance from any vector x, where $x \in X$, to the center m_i.

Each output node computes a weighted summation of the values generated by the hidden layer nodes, that is

$$y_j = \sum_{i=1}^{q} w_{ij}\psi i \tag{3}$$

where the w_{ij} values are termed the *weights*. Thus the output layer maps $\Psi \rightarrow Y$, Y being the space of all possible angular directions to the source, where $(y_1, y_2, ..., y_r) \in Y$ for r output nodes. Hence the term *basis* in radial basis function network indicates that the output lies in the linear space spanned by the outputs of the hidden layer nodes. Note that the output of the network for a single-source resolution network could be a single number indicating the angular direction to the source. However, since a future goal of this research is multiple-source detection and resolution, we designed our network so the output consists of a vector whose components represent angular "cells". The output values of these cells indicate the presence and strength of targets within those angular bins. The number of network output nodes therefore limits our potential angular resolution.

In keeping with the neural network paradigm, the determination of the *optimal* weights w_{ij} is by a modified gradient descent algorithm known as 'backpropagation'. The optimality criterion is a least mean square (LMS) error, where the error is the difference between the actual network outputs and the desired network outputs for all of the training samples. Each training sample consists of a measured input to the network for a given angle to a radiating source, and the corresponding desired network output for this angle. The radial basis function network is *adaptive*, that is it adds hidden layer nodes during the training phase whenever the network performance is unacceptable and there are no hidden layer node centers located in the function space 'close' to the current input. Whenever a hidden layer node is added, the center m_i is set equal to x, the vector value of the outputs of the input nodes for the input that yielded the unacceptable performance. The new weights w_{ij} for this node are set to values that force the network output to be within an acceptable interval about the desired output. Each σ_i for the new hidden node is set to an initial value which is either predetermined or else dependent upon the 'spacing' of the input vectors (see Section 4). Additional details regarding the network architecture and the adaptive training algorithm can be found in two articles by Lee [3, 4]. Although we followed the basic algorithms given in these articles closely, we also implemented variations based upon experimentation with our specific problem.

3 Weight Optimization Using LMS

The radial basis function neural network paradigm of Section 2 includes a modified gradient descent optimization of the *weights* (the w_{ij} of Equation 3). Unfortunately, it is just this weight optimization that introduces several vague aspects into the network design, such as requiring the network designer to choose parameters for which the literature only contains general guidelines. Two of the most important of these parameters include the descent *stepsize*, and the stopping criteria. The choice of the stepsize dictates the convergence speed of the weights and even whether or not they will converge or oscillate. One popular approach is to slowly change the value of the stepsize based on some monotonically decreasing function. This requires the choice of the initial stepsize value and the decreasing function. The weight changes must eventually stop, which is generally done when the network output error (the desired versus actual outputs for the training set) reaches a predetermined value, a value which must also be chosen by the network designer. And regardless of the choices for stepsize and stopping criteria, the gradient descent algorithm suffers from the possibility of stopping at a local, instead of global, minimum.

This section presents another approach to solving for these weights based on linear algebra techniques. Although the basic ideas have been previously presented [12], this paper gives new results regarding the choice of the most important network parameter, the σ_i from Equation 2. This linear algebra algorithm finds the globally optimum values for the network weights in the least mean square (LMS) error sense, using a matrix pseudo-inverse. Note that these weights will only be optimal with respect to whatever data is used to form the corresponding matrices (see below). This means that if the linear algebra approach is used as a direct replacement for the modified gradient descent weight optimization in the neural network, the weights will only be optimal for the *training* data, not any other data.

To convert the network to the linear algebra paradigm we change notation slightly from Section 2. Each input vector x_{lk} consists of n components, i.e. $k = 1, ..., n$; there are N input vectors, $l = 1, ..., N$. The N input vectors are the preprocessed data described in Section 2 computed from N phase measurements at N source angles. A subset, q, of the N input vectors is selected to *train* the neural network by solving for the weights. For example, if N were 121 (corresponding to a 1° data interval from −60° to +60°), we could train the network on the measured data at 10° intervals, in which case q would be 13. Data not used for training is used to evaluate the network's performance (see Section 3.1).

Before we discuss training, we define an $N \times q$ Gaussian function matrix Φ which includes all Gaussian function evaluations. The evaluation of each Gaussian function (the ψ_i of Equation 1) is denoted by ϕ_{li}, where l denotes the input vector number as above, and i denotes the Gaussian number, $i = 1, ..., q$ (i.e. there are q Gaussians). Each Gaussian's *center* (see Equation 1) is given by m_{ik}, with i denoting the the Gaussian number, and $k = 1, ..., n$ the vector component. The $N \times q$ Gaussian function matrix is

$$\begin{aligned} \Phi_{N,q} &= [\phi_{li}] \\ \Phi &= \left[\exp^{-\frac{\sum_{k=1}^{n} (x_{lk} - m_{ik})^2}{2\sigma_i^2}} \right] \end{aligned} \tag{4}$$

with $l = 1, ..., N$ and $i = 1, ..., q$.

For network training, in the example above, select rows $l = 1, 11, ..., 111, 121$ of Φ to form a 13×13 *training Gaussian function matrix*, Φ_t, which is used below in the LMS solution for the network weights. Next, however, we describe how the network output vectors are evaluated.

Output vector l has components y_{lj} for $l = 1, ..., N$ and $j = 1, ..., q$. The elements of the output matrix, Y, are weighted linear sums of Gaussian function values.

$$\begin{aligned} Y_{N,q} &= [y_{lj}] = [\sum_{i=1}^{q} \phi_{li} w_{ij}] \\ Y &= \Phi W \end{aligned} \tag{5}$$

The N rows of Y are the N output vectors. We know the desired outputs at the q training angles (−60°, −50°, ..., +50°, +60°). This corresponds to the rows of Y with $l = 1, 11, ..., 111, 121$ which we use to form a 13×13 desired output matrix Y_d. The w_{ij} are the weights from Equation 3. Note that W is a $q \times q$ matrix. If we have the weights, W, we can compute a predicted output vector for any input vector using Equation 5. Next, we describe the LMS solution for the optimal weights, W.

We solve for the LMS optimal weights based on minimizing the cost (error) function

$$(Y_d - \Phi_t W)^T (Y_d - \Phi_t W), \tag{6}$$

where $\Phi_t W$ approximates the desired output. The steps for computing the optimal W are:

1. Collect input data, i.e. the x_{lk}, for N measurements at N known source angles.
2. Use a subset, q in length, of the N input vectors to form a $q \times q$ matrix Φ_t. These q vectors are also used as centers m_{ik} (see Section 2).
3. Solve for W using the LMS solution which minimizes the cost function of Equation 6 [2], yielding
$$W = \left(\Phi_t^T \Phi_t\right)^{-1} \Phi_t^T Y_d. \tag{7}$$

The expression

$$\left(\Phi_t^T \Phi_t\right)^{-1} \Phi_t^T$$

is one of the left inverses of Φ_t. That $(\Phi_t^T \Phi_t)$ is invertible, yielding this unique solution to the LMS problem, is guaranteed by the linear independence of the columns of Φ_t [11]. In an upcoming paper we discuss the relationship between the linear independence of the columns of Φ_t and the specific physical configuration of the experiment. With respect to the results reported here, the Φ_t matrix always has linearly independent columns.

3.1 Backpropagation Versus Linear Algebra Weight Optimization

Theoretically, solving for the w_{ij} weights via the left pseudo inverse should yield network performance superior (for the training data) to that when the weights are found with the modified gradient descent algorithm, given that the performance is measured in terms of the least mean square (LMS). This implies that LINNET, the name of the network with the weights optimized using the pseudo inverse, should always outperform the neural network, whose weights are optimized using a modified gradient descent weight optimization, for a root mean square (RMS) performance criteria over the training data. Our empirical results verified this, given that the other network parameters (such as σ_i) are kept constant. This section addresses the question of how LINNET performance compares to the neural network for data that are not part of the training data. The name LINNET is inspired by it being a linear algebra based network.

Table 1 summarizes the performance of both the LINNET and NBF neural networks over ten different data sets. Each data set was collected within a short period of time during which the antenna's performance and configuration was assumed to remain constant. A data set contains the input vectors for source angles in one degree increments from −60° to +60°, for a total of 121 sample points. The training samples from each data set are those input vectors (and corresponding desired network outputs) at every 10 degrees, i.e. 13 of the 121 data points. Once the network weights are calculated, the network performance is computed over all 121 data points and reported in the table as RMS angle error in degrees. The RMS error is computed as

$$\sqrt{\frac{1}{121}\sum_{l=1}^{121}(\theta_l - \hat{\theta}_l)^2}$$

Data Set	Best σ for LINNET			Best σ for NBF		
	σ	LINNET RMS Error	NBF RMS Error	σ	LINNET RMS Error	NBF RMS Error
TEST97	0.800	0.41	1.06	0.640	0.50	0.53
TEST92	0.850	2.01	2.00	0.835	2.01	1.97
TEST87	0.800	1.46	1.42	0.760	1.46	1.41
TESTSC	0.750	1.67	1.72	0.705	1.67	1.59
DATA1	0.750	1.43	1.34	0.750	1.43	1.34
DATA2	1.000	14.98	15.24	0.810	15.45	15.06
TEST80	0.850	0.47	0.67	0.700	0.53	0.54
TEST95	0.750	1.37	1.31	0.735	1.37	1.29
TEST60	1.300	1.93	2.37	1.000	1.98	1.95
TEST59	0.800	0.47	0.88	0.665	0.52	0.56

Table 1: Comparison of NBF and LINNET RMS Error Performance

where θ_l is the source angle and $\hat{\theta}_l$ is the angle predicted by the network. The row labels are the names given to the different data sets. They represent different signal to noise ratios or minor changes in the physical configuration of the experiment, except that DATA2 was collected with coffee cans stacked in front of the array, and TESTSC incorporated a metal mesh screen mounted around the array to simulate an airframe superstructure. For the results reported in Table 1 all of the σ_i values are the same, i.e.

$$\sigma_i = \sigma_j \qquad\qquad i,j = 1,...,q$$

where q is the number of Gaussian functions. The columns under "Best σ for LINNET" show the RMS error for LINNET ("LINNET RMS Error") and the NBF neural network ("NBF RMS Error") with all the σ_i set equal to that value ("σ") which optimizes the RMS performance of LINNET. Similarly, the columns under "Best σ for NBF" report the RMS error performance with all the σ_i set equal to that value which optimizes the RMS performance of NBF. Note that the σ_i optimization is reaccomplished for each data set. Table 1 implies that the performance of the LINNET and the NBF neural network are very similar, especially when using the σ_i values optimized for the NBF, which is as expected.

3.2 Weight Sensitivity and Neural Network Training

One of the reasons for using the radial basis function network architecture is its reported order of magnitudes reduction in the number of iterations required for gradient descent weight optimization versus that required for *backprop* networks based on the sigmoid neuron function. Our results substantiated this trend, with the neural networks reported here achieving their performance after only three weight optimization iterations. With respect to these results we have found the probable reason for this rapid convergence, as shown in Figure 2.

The $W(3,3)$ and $W(12,13)$ axes represent weight values for the weights 'connecting' the output of the third Gaussian with the third output vector component, and the output of the

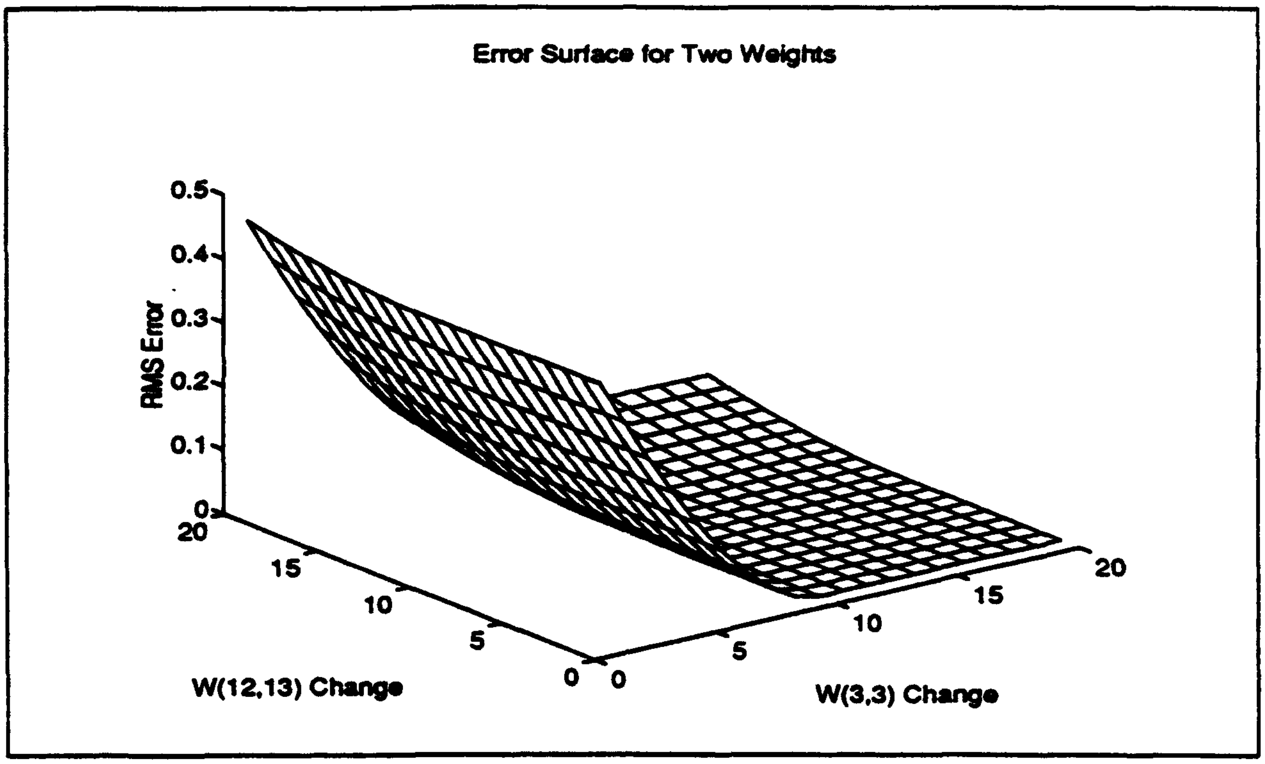


Figure 2: LINNET *Performance vs Weight Changes*

twelfth Gaussian with the thirteenth output vector component, respectively. The optimal values (found with the LMS algorithm) for each weight lie at the values denoted by 10 along the axes. Each unit value along the axes represent a change from the optimal value of ten percent. Thus a value of 20 denotes a 200 percent larger value than the optimum, while a value of 1 represents a ninety percent smaller value than the optimum (i.e. only one tenth of the optimal value).

Figure 2 is just one of many such plots created, and every one had the same basic shape, steep sides and a flat bottom. Although we didn't generate every possible plot (that would involve all possible pairings of the weights), we believe that the actual multidimensional error surface has the same basic 'shape'. Thus if the initial neural network random weights lie on a steep side, the modified gradient descent search should reach the bottom plateau relatively quickly. In addition, the flat bottom implies that weight values anywhere along this plateau should yield near optimum network performance. Thus the neural network would be expected to obtain near optimal performance in significantly fewer gradient descent iterations for this particular problem. than for many other types of problems, which is exactly what we discovered. This also implies that continuing the weight optimization would not lead to any significant increase in performance, also consistent with our empirical observations.

4 The Major Network Parameters

One major advantage of the linear algebra approach is that it does not require certain variables involved with the modified gradient descent weight optimization technique (such as the stepsize and stopping criteria) inherent to the neural network paradigm. However, both the neural network and linear algebra approach require determining values for two major network parameters, the Gaussian function 'centers' and the norm distance weighting (the Gaussian variances), which are respectively the m_{ij} and the σ_i (the root of the variance) from Equation 1. For both of these approaches the choice of the centers is dictated by the nature of the 'training' data; the centers are chosen to coincide with the location of the training samples in the input space X (see section 2 and [6]).

4.1 The σ_i Values

Consequently, the values of the σ_i remain as the major network parameters (disregarding the parameters associated with the backpropagation weight optimization). This section addresses two methods used in this research for determining σ_i values. The first is to simply set all σ_i equal to the same value. The 'best' value can be found by an optimization procedure. The second method is to set each σ_i independently, based on the Euclidean 'distance' between the Gaussian function centers (the m_{ik}).

The first method is the simpler of the two and also satisfies our requirement that computations for the Gaussian functions require only *local* data (for a detailed description of data locality see [6]). This means that the computation of each σ_i does not require any information that depends on the presence or absence of any other Gaussian functions except the ith. Since this method sets all σ_i values equal to a predetermined constant,

$$\sigma_i = \sigma_j \qquad\qquad i,j = 1,...,q$$

each σ_i can be computed independently of any other Gaussian function. Figure 3 shows how the NBF neural network RMS angle error varies as a function of σ_i (same value for all $i = 1,...,q$).

Experiments with LINNET also demonstrate the same type of relationship between the RMS angle error and the value of σ_i; a steep decrease in RMS error with increasing σ_i until an optimum value is reached, after which the RMS error only increases slightly for further increases in σ. To determine if this relationship varied markedly with different training set granularities, we repeated the NBF neural network training at every 5 degrees (instead of every 10) which yields 25 training points instead of 13. Figure 3 shows the results both for 10 degree (dotted line) and for 5 degree (solid line) training sets. Halving the granularity had no appreciable affect for this, nor for any other, of the data sets that we investigated. Note, however, that the optimum value of σ (i.e. minimum RMS error) is different for the two training intervals, 5° and 10°. Note also that these optimal values are different for various measurement sets, as shown in Table 1, although the relative shapes of the sigma-error curves are similar to those in Figure 3.

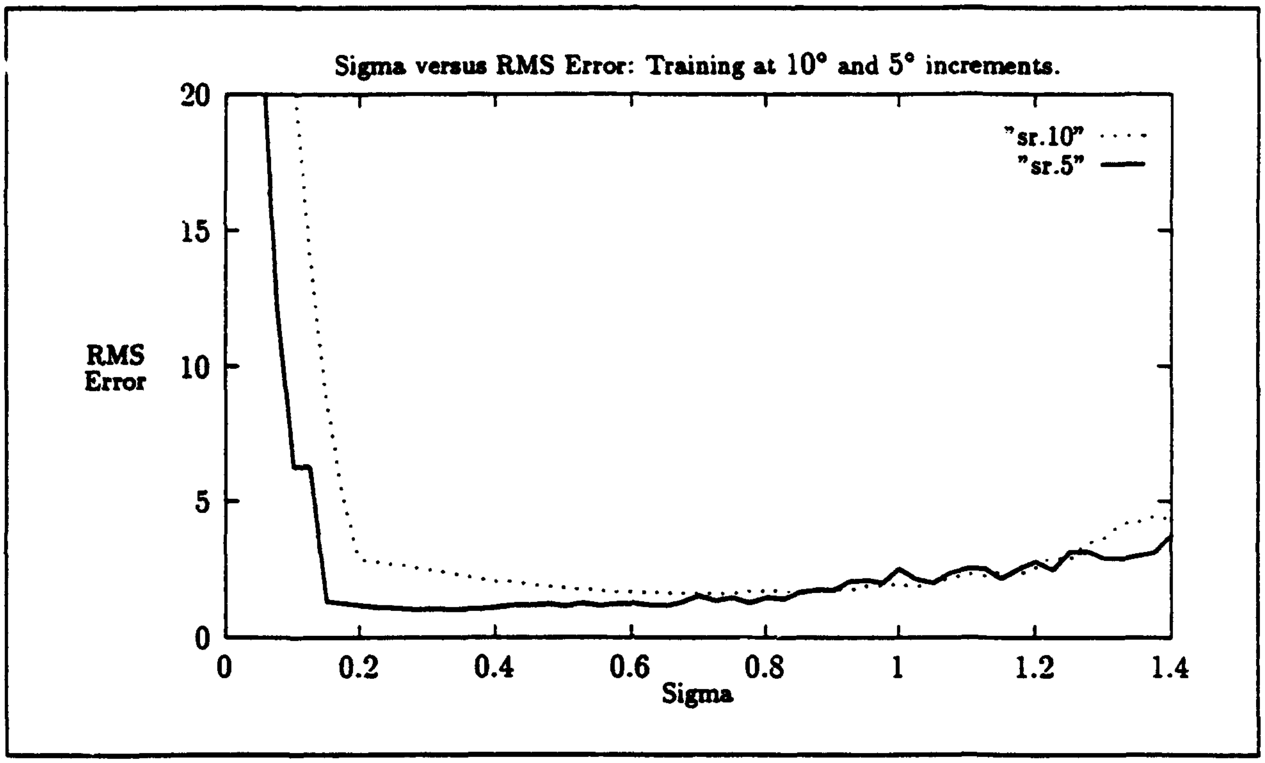


Figure 3: *RMS Network Error vs Constant σ_i Values*

The second approach uses different values for σ_i for each i, computed as a function of the Euclidean distance between neighboring Gaussian function centers (m_{ik}). Although this violates the locality requirement described above, this technique yielded a substantial increase in network performance for some of the data sets. A description of this approach follows.

For each 'new' Gaussian function i that is added to the hidden layer during training (see [6]), the Euclidean distance d_i to the last previous Gaussian function $i-1$ is computed by

$$d_i = \sqrt{\sum_{k=1}^{n} |m_{i,k} - m_{(i-1),k}|^2} \qquad i = 2, \ldots, q$$

with d_1 set equal to a predetermined constant. One analysis has shown that the scaling factor in the weighted norm of Equation 2 should be proportional to the distance between the centers of the Gaussian functions [5], i.e.

$$\frac{1}{2\sigma_i^2} = dis \cdot d_i$$

where *dis* denotes the proportionality constant, and theoretically, $dis \leq 1$. Thus one formulation for computing σ_i is

$$\sigma_i = \frac{1}{\sqrt{2 \cdot dis \cdot d_i}}. \tag{8}$$

This technique yields a different σ_i for each Gaussian function, yet requires setting only one parameter, *dis*. Just as with the first approach that used one value for all σ_i, the results for this approach reflect computer assisted searching for the optimum value of *dis*.

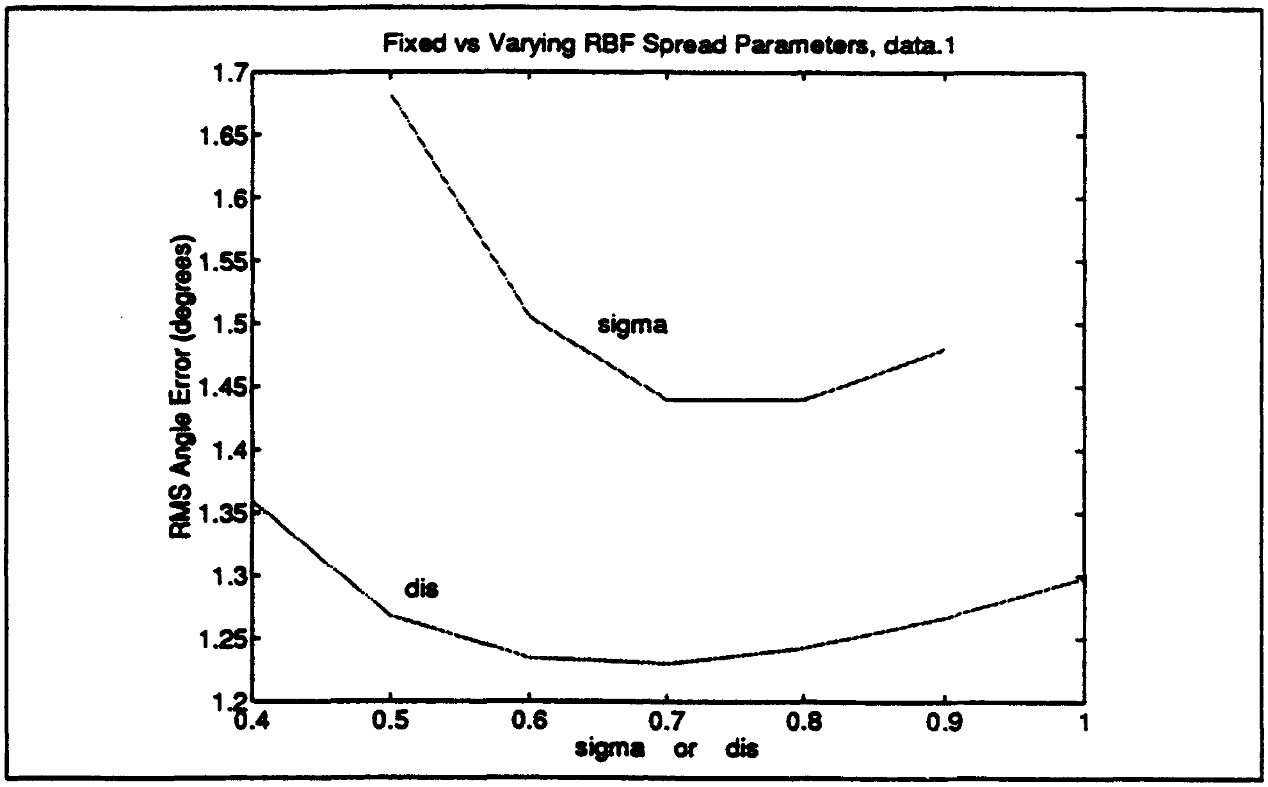


Figure 4: *Comparing Methods for Optimizing* σ_i *for* DATA1

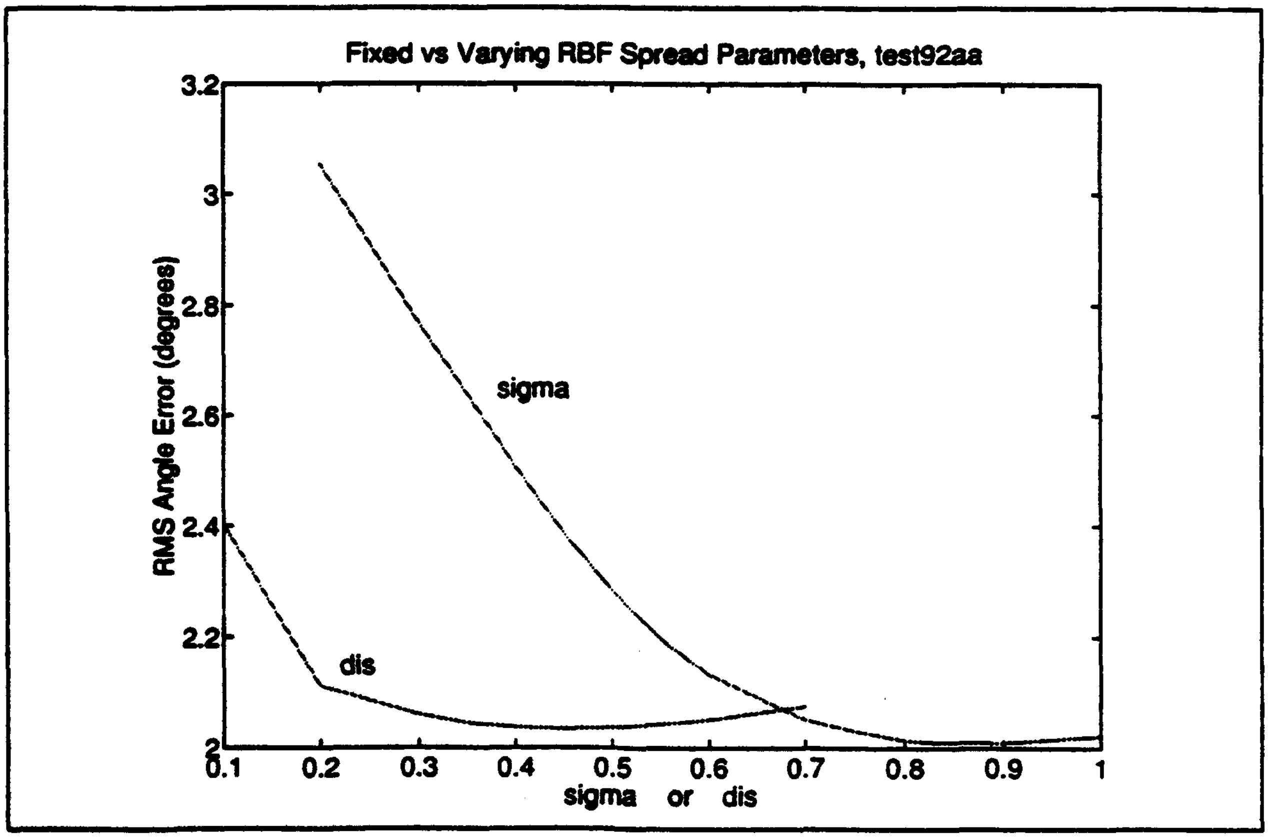


Figure 5: *Comparing Methods for Optimizing* σ_i *for* TEST92AA

Figures 4 and 5 show the LINNET RMS angle error performance with respect to both of these approaches for two of the data sets. In each figure the curve labeled "sigma" shows the LINNET performance with all of the σ_i values set equal to a constant. Note that this curve can be directly compared to the neural network performance (same technique for setting σ_i) from Figure 3. These two curves (labeled "sigma") show the same basic shape as that from Figure 3. The two curves labeled "dis" show the LINNET performance using the σ_i computations from Equation 8. For the DATA1 data set this second approach (using *dis*) yields approximately fifteen percent improvement, while for the TEST92AA data set, the first approach is slightly better. These two plots are included because they represent the two ends of the spectrum. For the other data sets the use of Equation 8 yielded something between zero and fifteen percent improvement. Ongoing research is still trying to determine what physical characteristics influence this spread of performance improvements.

5 Conclusion

We developed a new network, LINNET, which solved exactly for the output weights to minimize the network's RMS-error across a training set. We compared this network to NBF, which uses a gradient descent technique which minimizes the error by converging these parameters. Our comparisons of these two networks verify our hypothesis that NBF is finding an acceptable near-global minimum after very few training iterations. We presented a typical "bowl-shaped" tomography of an error-weight surface which indicates why this is not surprising. Finally, we presented two techniques which were tested to determine an optimal network σ. We determined empirically that the better choice of technique varies between data sets. An area of ongoing research is the investigation of physical parameters of the antenna array and environment which influence this performance improvement.

Acknowledgments

This work was sponsored by the Air Force Office of Scientific Research as part of the initiative on "Adaptive Optics and Beamforming." Support by ARCON Corporation was provided under USAF Contract #F19628-92-0177.

References

[1] D. S. Broomhead and D. Lowe. Multivariable Function Interpolation and Adaptive Networks. *Complex Systems*, 2:321–355, 1988.

[2] B. D. Carlson and D. Willner. Antenna Pattern Synthesis Using Weighted Least Squares. *IEE Proceedings - H*, 139(1):11–16, February 1992.

[3] S. Lee. Supervised Learning with Gaussian Potentials. In B. Kosko, editor, *Neural Networks for Signal Processing*, pages 189–227. Prentice Hall, Englewood Cliffs, NJ. 1992.

[4] S. Lee and R. M. Kil. Bidirectional Continuous Associator Based on Gaussian Potential Function Network. In *Proceedings of the 1989 IEEE IJCNN Volume I*, pages 45–53, San Diego, CA, 1989. IEEE.

[5] F.J. Narcowich, N. Sivakumar, and J.D. Ward. On Condition Numbers Associated with Radial-function Interpolation. Technical Report CAT Report 287, Center for Approximation Theory, Dept of Mathematics, Texas A&M University, January, 1993.

[6] T. O'Donnell, J. Simmers, and D. J. Jacavanco. Neural Beamforming for Phased Array Antennas. In *Proceedings of the 1992 Antenna Applications Symposium (Robert Allerton Park)*, Griffiss AFB, NY, September, 1992. USAF HQ Rome Laboratory.

[7] T. Poggio and F. Girosi. Networks for Approximation and Learning. *Proceedings of the IEEE*, 78(9):1481–1496, September 1990.

[8] M. J. D. Powell. Radial Basis Functions for Multivariable Interpolation: A Review. In J. C. Mason and M. G. Cox, editors, *Algorithms for Approximation*, pages 143–167. Clarendon Press, Oxford, 1987.

[9] J. A. Simmers and T. O'Donnell. Adaptive RBF Neural Beamforming. In *Proceedings of the 1992 IEEE Mohawk Valley Section Command, Control, Communications, and Intelligence (C3I) Technology & Applications Conference*, pages 94–98. IEEE, June, 1992.

[10] D. F. Specht. A General Regression Neural Network. *IEEE Transactions on Neural Networks*, 2(6):568–576, November 1991.

[11] G. Strang. *Linear Algebra and its Applications.* Academic Press, New York, 1980.

[12] A. C. Tsoi. Multilayer Perceptron Trained Using Radial Basis Functions. *Electronics Letters*, 25(19):1296–1297, September 14 1989.

HIGH PERFORMANCE MICROWAVE NEURAL NET-ANTENNA ARRAYS FOR RADAR SIGNAL CLUSTERING

A. D. Macedo Filho [ø⚓] and **H. D. Griffiths** [ø]

[ø] **University College London, UK**
[⚓] **Brazilian Navy**

ABSTRACT

Microwave neural networks are introduced as radar signal clustering networks for applications in ESM systems. Some early designs and their associated problems are then explained. Finally, a new kind of structure is proposed, making use of the signal phase in the classification process. The results provided by some simulated topologies are shown and discussed.

1. INTRODUCTION

Modern ESM processors must provide an effective interception performance in today's increasingly dense and complex signal environments. Moreover, they must provide accurate analysis in less than 1 second [1]. Hence, fast signal clustering and classification represent demanding needs.

Antenna arrays are suitable for fast signal processing, either by using distributed digital networks or directly by means of microwave structures such as Rotman Lenses or Butler Matrices [2]. Furthermore, adaptive antenna arrays [3] may be viewed as a form of "intelligent" microwave system. They can respond to an unknown signal environment by processing the array outputs adaptively, attempting to optimize the signal-to-noise ratio at the front-end or receiver output. Their intelligent behaviour consists of producing the appropriate problem-solving response when faced with a particular problem stimulus. However, their skill is limited to a very specific task. Similar systems should be able to associate responses to stimuli, and hence to generalize - that is, small input changes produces small output changes. Such behaviour is exactly that of "clustering" or "classification".

Therefore, the original adaptive arrays may be evolved to recognize and classify emissions. As a matter of fact, one of the most primitive and common activities of animals (man included) is sorting like items into groups. These groups are defined by patterns which are recognized, thus enabling classification. In broad sense, this is fundamentally the platonic philosophy of perfect structures, in which there is a set of "ideal forms" from those the real world present the observer imperfect replicas. In psychology, pattern recognition is defined as a process by which "... external signals are converted into meaningful perceptual experiences" [4].

Adaptive arrays, in spite the fact that they can learn, are unable to maintain the acquired knowledge as the adaptation process is endless. They are continuously learning from new data and discarding previous information. Obviously, while presenting such behaviour, such a system is unable to undecode the information stored in its synaptic weights.

Essentially, there are two main phases in intelligent behaviour. The first one is the learning phase, in which some specific ability is slowly acquired. The second phase occurs when the skills that have been learned are put into practice, which is referred as application phase [5]. Microwave neural systems conform to this two step framework.

The main objective of microwave neural nets in this application is to provide radar signal clustering for ESM purposes, using microwave hardware. This should improve system integration and enable the received and analysed radar pulses to be directly sent to the ECM subsystem without further need of re-constructing detected signals in real time.

2. ANTENNA ELEMENTS AS SENSORIAL DEVICES

In the scene analysis discussed in the previous section, one can define the pattern space as the domain of data observed by the sensory organs or devices.

One way to characterize a signal s is to assign it to a range of finite measurable relevant parameters. The column vector x in such space is

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{1}$$

where $x_i = f_i(s)$, and $\{f_i\}_{i=1}^{n}$ is a set of measurement procedures, or primitives, into which s is decomposed.

These primitives can be of two types. They are either parameter specialized, when each sensitive organ or measuring device is able to work selectively over one unique specific parameter; or generic, in which case all parameters are processed at once, but in different ways, by all measuring devices.

Most ESM systems detect the incoming signal and then obtain parameter specialized primitives. The approach suggested here is, however, towards generic primitives that are obtained from inhomogeneous antenna arrays. Naturally, the more inhomogeneous the array (that is, the more its elements differ from each other), the more it is able to react differently to signals of different frequency, polarisation, bearing, etc... Therefore, by presenting better sensitivity the network has improved conditions to cluster the incoming signals into well separated subsets. Such clustering may not be viewed as being "right" or "wrong" since there are many possible criteria that may be adopted. A set of clusters will not be unique, and new insights and understanding can arise from clustering results, thus unveiling unexpected solutions. Nevertheless, multiple classification procedures are able to minimise the errors in a desired classification, once combinations of different results are explored.

Usually, in order to cluster the incoming signals, the network will have to learn from a set of examples. This is cognitive system operation. Microwave networks, however, are different from conventional neural systems as they must base their capabilities on the phase response rather than the amplitude response, being so a very particular class of neural systems.

3. INITIAL MICROWAVE NEURAL APPROACHES

The first microwave system that we considered used signal amplitudes as the basis of its classification [6]. It was a typical "back-propagation" network [7] where the first layer had enhanced processing capability since it could deal with the complex stimuli output by the antenna array.

The delta-rule, LMS (root-mean-square), or Widrow-Hoff learning method, implicit to the back-propagation paradigm, calculates an error value given by

$$\varepsilon_j = d_j - y_j \qquad (2)$$

where

d_j is the desired response, and

y_j is the desired output, expressed by

$$y_j = g\left(\sum_{i=1}^{n} w_{ij}X_j + L_j\right) \qquad (3)$$

where g(.) is a generic nonlinear function (usually a sigimoidal curve);

X_j is the input to neuron j;

w_{ij} is the synaptic weight value from neuron i to neuron j; and

L_j is the threshold value of neuron j.

The adjustments of each w_{ij} (j,i = 1,2,....n) is given by

$$w_{ij,\,t+1} = w_{ij,\,t} + \mu\varepsilon_j X_j$$

where μ is a positive scalar, $0 < \mu < 1$, used to control the learning rate.

In the proposed system the first layer performed complex weighting and sum, but outputs only scalar values. It works as if the incoming microwave signal were detected just after the combiner by a non-linear device. Thus, in spite of the fact that the variables X_j and W_j (the synaptic weight vector formed by the several individual weights w_{ij}) are complex, y_j, d_j and ε_j are real scalar values. In fact, this complex layer is an interface that conforms the microwave signals to the rest of the system, a regular back-propagation network as shown in Fig 1.

The activation function g(.) for this first layer was defined as

$$g(x) = \frac{x^2}{1 + x^2} \qquad (5)$$

This function is close to the typical sigmoid, and in fact corresponds to the operational curve of microwave detectors in the small-signal regime, where x << 1. Table I shows the differences between the two types of layers applied.

The three antennas considered for the simulations were of different polarisations: vertical, horizontal and left-hand circular (LHC). The gain of all these elements were considered constant over all the frequency band. This will not be true in practice, of course, and such a constraint was adopted only to simplify calculations. Two signal parameters were chosen as being of particular interest: frequency and polarization.

The training set comprised three signals: one 9 GHz LHC considered as a threat, and two friendly emissions, one at 8 GHz horizontally polarized and the other at 10 GHz 45° slant polarized.

Fig 2 shows the progress in performance of this network along the learning phase. The generalization capability is presented in Fig 3.

These results were rather promising; nevertheless a problem arises when one has to consider that the incoming amplitudes are completely beyond the system's control. It is not possible, in practice, to determine the incoming signal's amplitude in advance, and so the training process is simply prohibitive with such a wide range of possibilities. Another difficulty is that if video layers are introduced there is no advantage in terms of fast ECM reaction.

A second system was further investigated in the laboratory [8] in order to avoid these practical restrictions. The clustering process should now be carried out in terms of the output phase rather than in the amplitude.

This simple apparatus is shown in Fig 4a, and it had phase-weights only, in order to be easily assembled. Two classes of signals were established. Class A comprised a 11 GHz vertically-polarized and 12 GHz horizontally-polarized signals. Class B contained only a 12 GHz vertically-polarized signal. This system's responses are indicated in Fig 4b. It is verified that the two classes were successfully clustered, each one in one semi-circle after only 10 trial-and-error weight settings. These responses were shown as well to be robust to 100 MHz variations and to 30° tilts in polarisation.

4. HIGH PERFORMANCE MICROWAVE NEURAL SYSTEMS.

Recent investigations delineated the complex Widrow-Hoff method [9] to realise the learning routine to such complex phase-networks. This rule, although originally intended to amplitude based systems, has no restrictions in terms of phase and so is well suited to the present case. The adjustment procedure is now defined by:

$$w_{j,\,t+1} = w_{j,\,t} + 2\mu\varepsilon_j X_j^* \qquad (6)$$

where the symbol * denotes the complex conjugate, and all variables are assumed to be complex except for μ, which is a real scalar.

In this manner, the generic microwave neuron is pictured in Fig 5.

Since the clustering information is carried on the output phase, there is no need to place a nonlinearity after the linear combiner. This follows because unequal amplitudes are expected at the front end of the combiner; consequently the final phase is not the sum of the individual phases. Thus, this process by the phase viewpoint is intrinsically non-linear. A second advantage in avoiding non-linearities is that this avoids the possibility of intermodulation.

Some possible architectures for these systems are illustrated in Fig 6. The emissions are considered to come from boresight. Extra phase shifters could phase-steer the sight angle of the array if necessary.

The simulations were made considering a frequency range of -300 MHz to +300 MHz around the central frequency of the training signal.

Fig 7 presents the results for a simulated single neuron trained to separate one 8 GHz horizontally-polarized signal from a 9 GHz vertically-polarized and a 10 GHz horizontally-polarized signals. A second clustering could be defined separating now the 9 GHz signal from the other two. It may be seen that if both responses are analysed, then each signal could be pinpointed.

The response for a more challenging training set comprising a 9 GHz vertically-polarized and a 10 GHz horizontally-polarized signals considered Class C, and a 9 GHz horizontally-polarized signal as being Class D, is presented in Fig 8.

A more complex network which we have named "the Fish" is shown in Fig 9 to cluster now 3 signals as belonging to Class A (the same two in Class C plus an 8 GHz 45° slant polarized signal) from two defined as Class B (comprising class D plus a 10 GHz LHC polarized signal). The results for the output neuron (f) are provided as well as for one of the neurons of the second layer (u). It is interesting to note that (u) has a better performance than (f) after these 300 training loops. The redundancy is quite interesting for reliability purposes.

5. CONCLUSIONS

Microwave Neural systems have been studied, and appear to provide a possible alternative to conventional techniques in performing radar signal clustering in real time. The results obtained through MATLAB [10] confirmed that these systems are potentially useful for ESM purposes. Nevertheless, there are yet several points to be established to realise in practice such networks with high degree of confidence, such as:

a) the responses of phase shifters and attenuators/amplifiers are not independent of frequency, and so their particular responses may need to be included on the training;
b) the precision of the phase shifters must be examined;
c) the use of lensoids, that is, lens-like structures with suitable edges shaped not to provide beam-forming, but clustering-forming may be studied as well; and
c) the possibility to perform unsupervised learning may need further investigation.

6. ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Prof R. Benjamin and to Prof K. Milne for the invaluable advises. Mr Macedo is sponsored by the Brazilian Navy.

7. REFERENCES

1. Cussons, S, Roe, J. and Feltham, A. (1990) Knowledge based signal processing for radar ESM systems. Paper presented at Military Microwaves Conference, London, UK.
2. Archer, D.H. (1984) Lens fed multiple beam arrays, *Microwave J.* vol 27 n.9:171-195
3. Hudson, J.E. (1981) Adaptive array principles. Peter Peregrinus Ltd., London U.K.
4. Lindsay. P.H. and Norman, D.A. (1972) Human information processing. Academic Press, N.York, USA
5. Kandel, A. and Byatt, W.J. (1978) Fuzzy sets, fuzzy algebra and fuzzy statistics, *Proc. IEEE* 66 (no.12):1619-1631

6 Macedo F, A.D. and Cardoso, L.A.L.S (1990), Microwave neural net-antenna array for electronic warfare systems. Paper presented at the European Microwave Conference, Budapest, Hungary

7 Simpson, P.K. (1988) Artificial neural systems, Pergamon Press, N.York, USA

8 Macedo F, A.D. and Mosso, M.M. (1991) Microwave neuron: an insight in sensory-like signal processing applying microwave devices. Paper presented at the SBMO Conference, Rio de Janeiro Brazil

9 Widrow, B., McCool, J. and Ball, M. (1975) The complex LMS algorithm, *Proc. IEEE* 63 :719-720

10 The MathWorks, Inc. (1989) MATLAB Users Guide

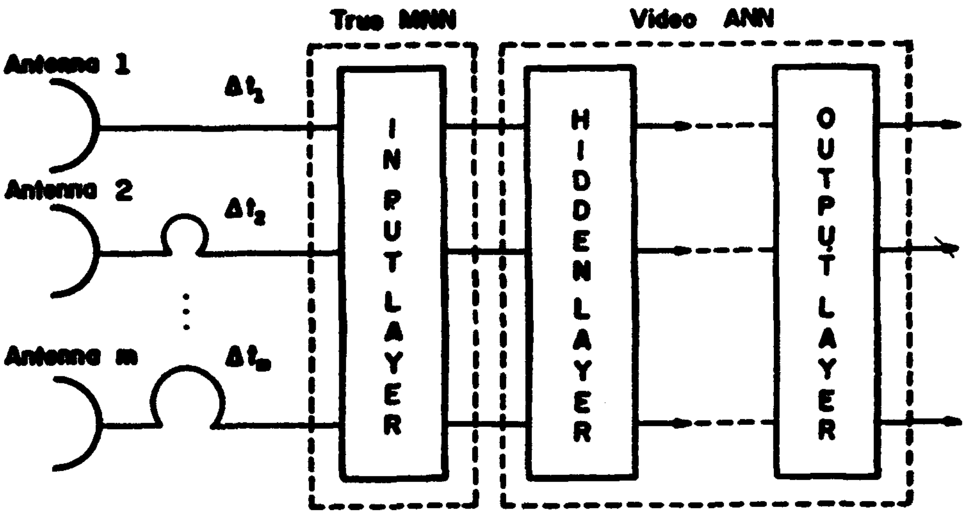

True MNN
Video ANN
Antenna 1
Antenna 2
Antenna m
Δt1
Δt2
Δtm
INPUT LAYER
HIDDEN LAYER
OUTPUT LAYER

Fig 1 - Early Amplitude Driven Microwave Neural-Network

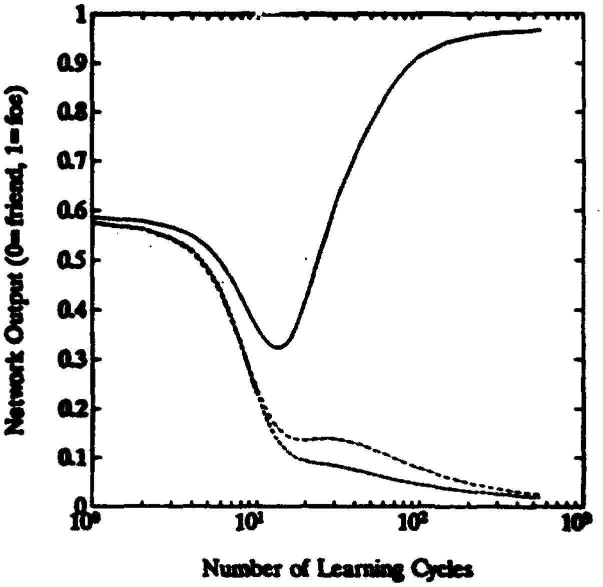


Fig 2- Evolution of the friend-or-foe detection abilities of the network, due to the learning process in the microwave layer. The solid line is the response of the network towards a "threatening" 9 GHz, counterclockwise circularly polarized signal. The dashed line and the dash-dotted line are the responses of the network towards two non-theatening signals, one 8 GHz, 0 degrees linear polarization and the other 10 GHz, 45 degrees linear polarization.

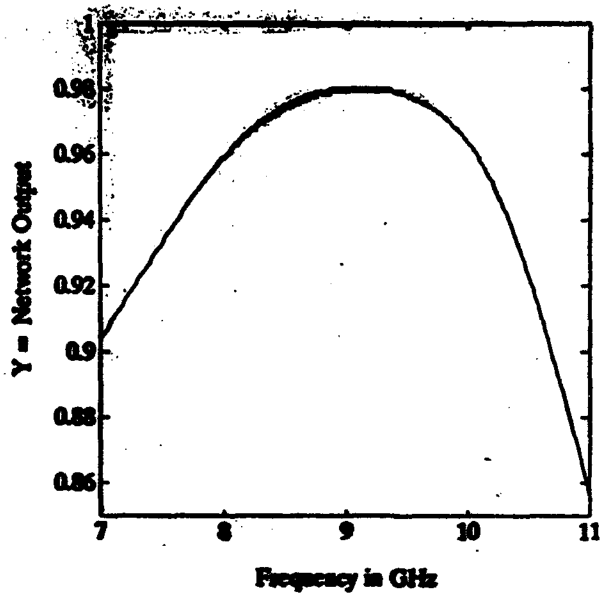


Fig 3- Network output as a function of frequency, for a counterclockwise circularly polarized signal. Note that the network remains exhibiting output values close to one (foe) for frequencies in the vicinity of 9 GHz, the "threat".

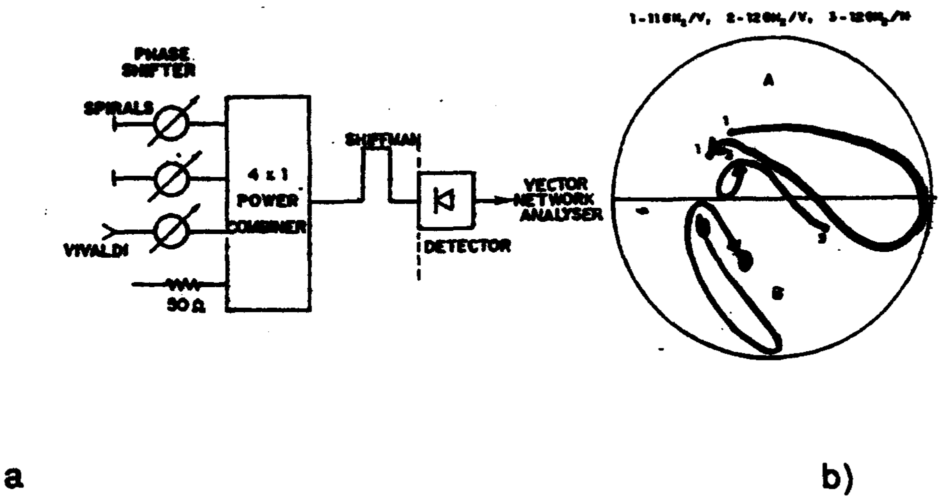

PHASE SHIFTER
SPIRALS
VIVALDI
50 Ω
4 x 1 POWER COMBINER
SHIFFMAN
DETECTOR
VECTOR NETWORK ANALYSER
1-11GHz/V, 2-12GHz/V, 3-12GHz/H
A
B
1
1
2
3
a
b)

Fig 4 Experimental Phase Neuron Prototype

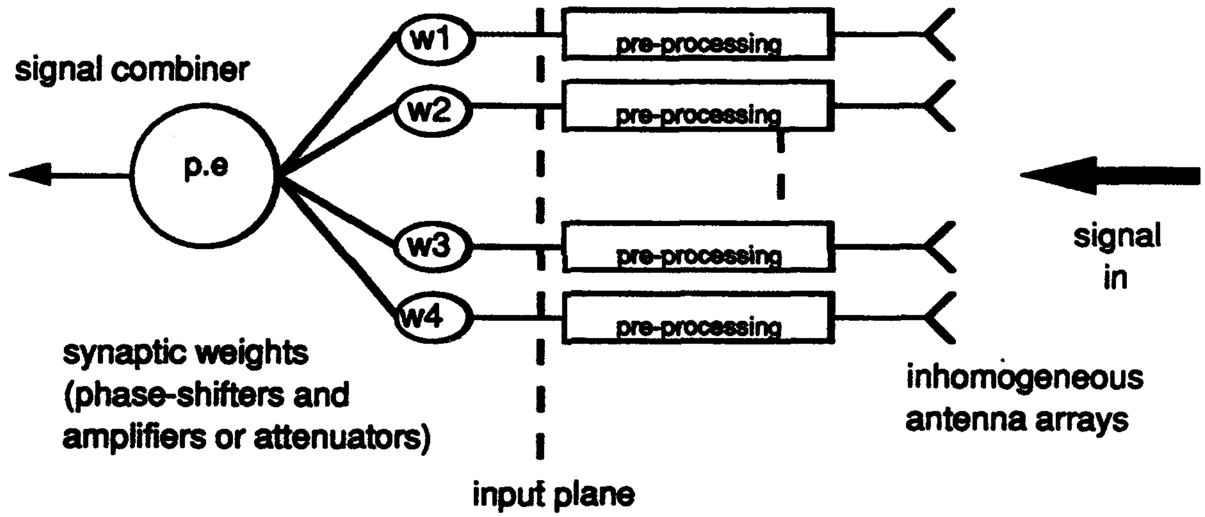

signal combiner
p.e
w1
w2
w3
w4
pre-processing
pre-processing
pre-processing
pre-processing
signal in
synaptic weights (phase-shifters and amplifiers or attenuators)
input plane
inhomogeneous antenna arrays

Fig 5- Generic Microwave Neuron

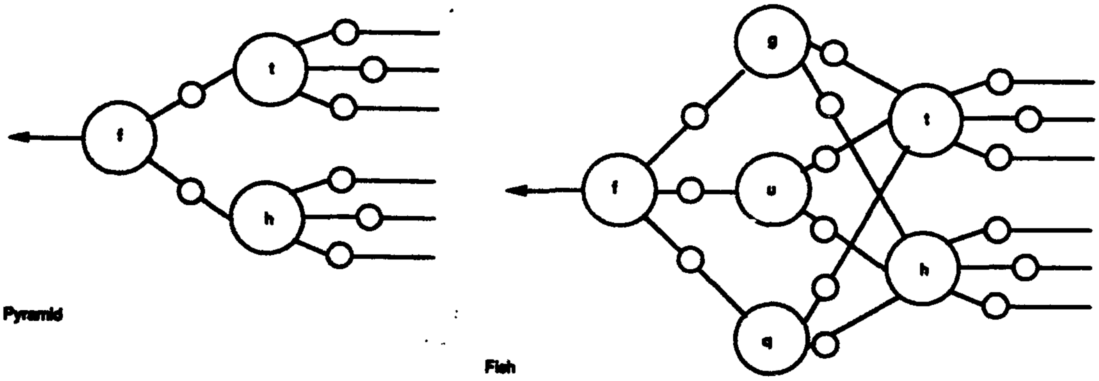
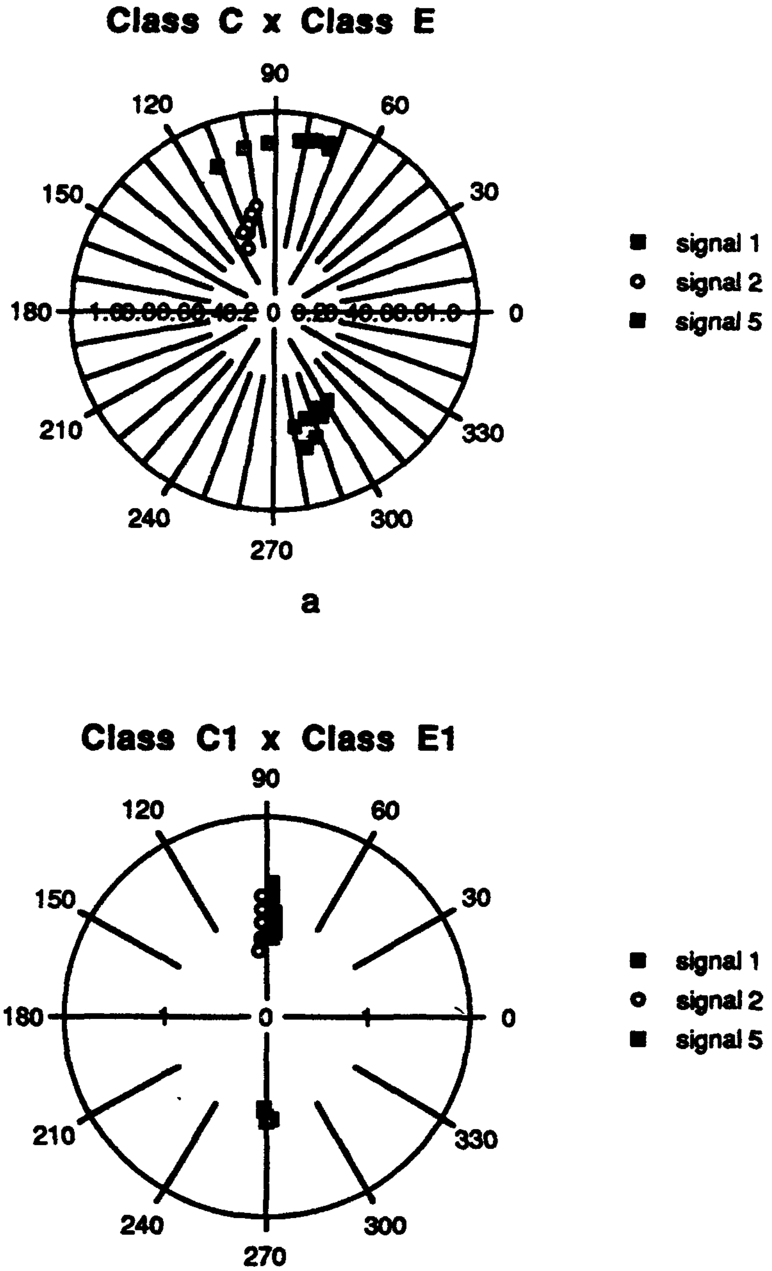

f
t
h
Pyramid
f
g
u
q
t
h
Fish

Fig 6- Some Possible Microwave Neural Topologies



a

b)

Fig 7- Results for a single neuron cosidering a)two signals in Class C (9Ghz vert-pol and a 10 Ghz horiz-pol) and one in Class E (8 Ghz horiz-pol);and b)two signals in Class C1 (8Ghz horiz-pol and 10Ghz horiz-pol) and one in Class E1 (9Ghz vert-pol)

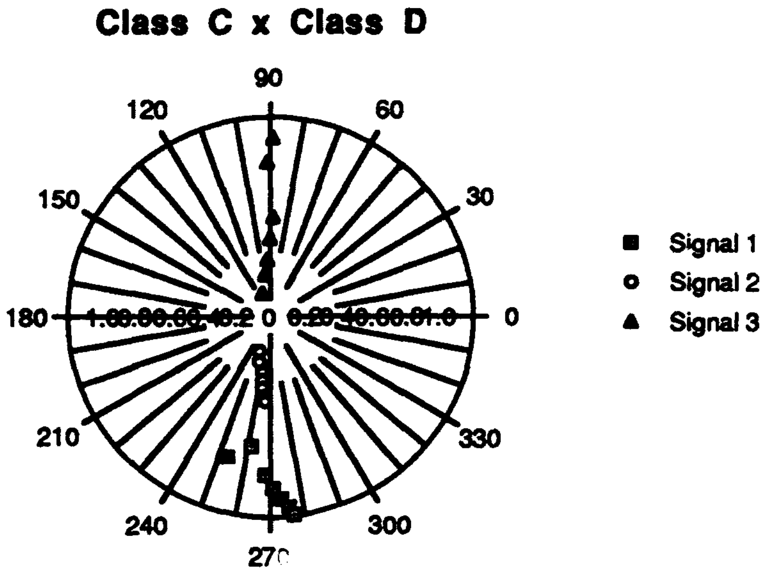
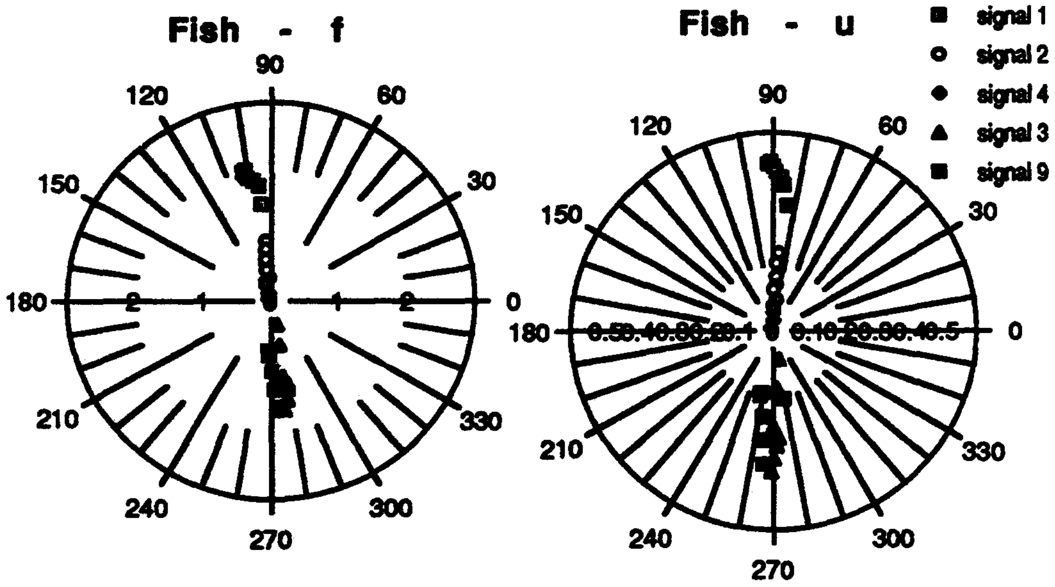

Class C x Class D

Fig 8- Results for a single neuron cosidering two signals in Class C (9 Ghz vert-pol and 10 Ghz horiz-pol) and one in class D (9Ghz horiz pol)


Fish - f

Fish - u

Fig 9- Response for Fish Network provided by neuron f and by neuron u

A NEURAL NETWORK ALGORITHM FOR IMPROVING THE PERFORMANCE OF DISTORTED REFLECTOR ANTENNAS

R. J. Bastian, W. T. Smith
Department of Electrical Engineering
University of Kentucky
Lexington, Kentucky 40506-0046

Abstract

Reflector antennas need to be electrically large to achieve the high gain, low side lobe performance required for various remote sensing and space communications applications. These large reflectors are susceptible to distortions in the reflector surface. The distortions in the surface are due to gravitational and thermal forces and structural design details such as support ribs and slots between panels. Active mechanical adjustment of the surface can compensate for the larger surface displacements. At higher frequencies, the subtle residual errors can remain electrically significant. Electromagnetic compensation may be used to supplement the mechanical techniques. In this paper, an electromagnetic surface error compensation technique using a neural network computing algorithm is presented.

1.0 Introduction

Adaptive reflector surface technologies can mechanically compensate for larger space reflector antenna surface distortions caused by gravitational and thermal forces. Electromagnetic compensation for surface errors can be used to supplement mechanical techniques at frequencies where subtle residual errors in the reflector surface remain electrically significant. Electromagnetic compensation techniques are implemented with algorithms that control the amplitude and phase of array feed elements.

The constrained least squares (CLS) compensation method matches the actual aperture field of a distorted reflector to some desired aperture field distribution at discrete points[1]. In a previous study[2], the CLS compensation method was found to be relatively insensitive to errors in the amplitude and phase of the complex excitations. This robustness prompted further investigation to determine whether neural network computing could be used to approximate the compensation performance of the CLS technique[3].

The advantage to using the neural network algorithm is that, once trained, the large computational overhead associated with the CLS would be relieved, thus facilitating near real-time control of surface error induced pattern effects. Training the networks can be very burdensome. Decomposition of the surface error results in an approximate superposition of excitations which relieves a great deal of the computational burden associated with the training. Brief summaries of reflector surface error effects, electromagnetic compensation techniques and neural networks are given in the next two sections. The issues concerning the training of the networks are summarized and a neural network compensation algorithm is outlined in Sect. 4. Example compensation calculations are presented in Section 5. The study is summarized in the last section.

2.0 Reflector Surface Errors and Compensation Techniques

In this section, the effects of surface errors and existing electromagnetic compensation methods are summarized. The study of reflector surface profile errors is generally separated into two classes: random and deterministic surface errors. Random surface error analysis generally concerns reflectors with rough surfaces. The effects of rough surfaces are usually analyzed using statistical methods. Deterministic surface errors arise from structural design details (ribs, slots, etc.) or distortions due to gravitational or thermal forces.

Random surface errors have been the topic of several investigations. The average effects of random surface errors and how rough surfaces degrade the antenna performance are well understood. The effects of random surface errors cannot be compensated for in practice due to the resulting wide field distribution across the focal plane. Hence, attempting to compensate for the roughness would require an unrealistically large feed array.

Slowly varying deterministic surface errors cause interference in the radiation pattern near the main beam and also in the fields near the reflector focus. Errors that exhibit periodic behavior are often more destructive as they can cause significant grating-type lobes in the radiation pattern and the focal plane field distribution[4]. It has been shown that periodic errors with increasing spatial frequency across the aperture cause distortions in the radiation pattern farther from the main beam than the lower spatial frequency (slowly varying) errors[5]. The slowly varying errors also cause distortions in the focal plane field distribution near the focus. Therefore, these slowly varying-types of errors can be compensated for using a moderate size feed array.

Electromagnetic (EM) compensation for surface errors is accomplished by using an array feed to capture unfocused energy around the focal point and adding the individual feed signals in a coherent manner. There are several techniques that have been used to perform this coherent summing of the signals.

Rudge and Davies[4] wrote one of the most referenced papers for electromagnetic compensation for surface errors. Their compensation approach exploits the Fourier transform relationship between the aperture field distribution and the focal plane fields on reception. A second Fourier transform is performed using a Butler matrix feed[4]. The initial hardware implementation was restricted to a cylindrical reflector system[4]. In their conclusions, the authors cited practical problems for implementing the technique for two-dimensional problems[4].

Conjugate field match (CFM) for reflector antennas has been widely used by many investigators for both reflector scanning and electromagnetic compensation. The technique attempts to maximize the power received from or transmitted in a given direction. The CFM excitation coefficients for an array feed are proportional to the complex conjugates of the received focal plane fields of each feed element when the reflector is illuminated by a plane wave[6].

Bailey[7] developed a method for computing the array feed excitations that modify the uncompensated distorted reflector fields such that they approximate the desired fields of a smooth reflector at discrete points in the aperture plane. The fields are approximated in a least squares sense[7]. It was discovered that while the pattern improved, the method could cause unacceptably high spillover loss. This was remedied by adding a constraint to the solution that limited the spillover power to reasonable levels[1]. This is known as the constrained least squares (CLS) technique.

The CLS compensation technique is very computationally intensive. Computing the array feed excitations can take several hours. The results of a previous study showed that the CLS method is fairly insensitive to errors in the amplitudes and phases of the complex excitations[2]. In this study, the feasibility of approximating the CLS technique with trained neural networks is demonstrated. Using the neural networks to compute the compensation excitations greatly reduces the computing time and facilitates near real–time control of the surface error effects. Discussion of the neural networks used in this study is presented in the next section.

3.0 Multiple Layer Feed–Forward Neural Networks

Neural networks are highly parallel, interconnected networks of simple computing elements loosely modeled after neurobiological systems. The most popular applications of neural networks are in the field of pattern recognition. The basic computing elements are step functions or other nonlinear functions. The elements are called nodes and are interconnected by weights that are adapted through a training process. Neural network types are defined by the network structure, nodal characteristics and the training or learning rules.

Multiple layer feed-forward networks are used in this study. Feed-forward networks are trained to provide a mapping from a given input space to a desired output response. Given a number of input-output pairs, the multiple layer feed-forward networks can be trained to learn some underlying function.

The structure of a multiple layer feed-forward network is shown in Fig. 1. The network consists of a set of input nodes which are typically interconnected to one or more hidden layers of nodes and eventually to a set of output nodes. The interconnections are weighted. With the exception of the input nodes, the output of each node is the weighted sum of the previous layer node values passed through the nonlinear function. The nonlinear function used in multiple layer feed-forward networks is the sigmoid function. The sigmoid nonlinearity allows the network to approximate a continuous valued function. An example sigmoid transfer characteristic is shown in Fig. 2.

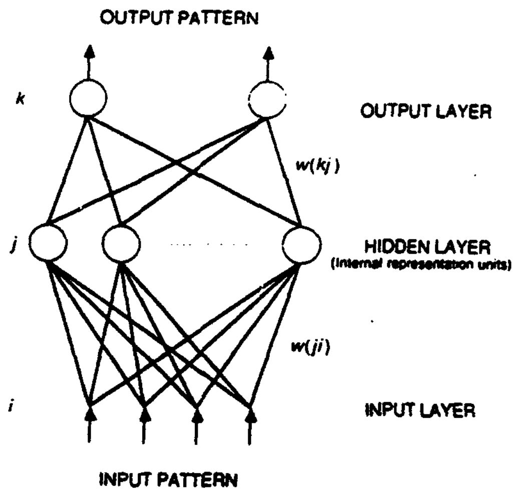


Figure 1. *A schematic depiction of a multiple layer feed-forward network. This example has only one hidden layer.*[8]

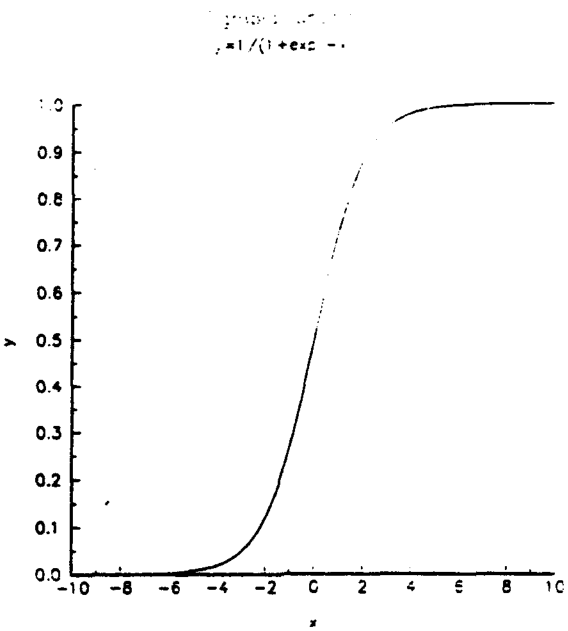


Figure 2. *An example of a sigmoid function transfer characteristic. The output of the sigmoid function is bound between 0 and 1.*

The network is trained using the backpropagation training algorithm. A number of known input–output pairs are presented to the network. The weights are initially set to small random numbers. The values for the known inputs propagate through the network. In general, the outputs *{o_k}* will not match the desired known outputs *{t_k}*. The error at the output is given by[8]

$$E = \frac{1}{2}\sum_{k}(t_k - o_k)^2 \tag{1}$$

The weights are then modified using a gradient algorithm to try to minimize the error. The training process is continued with the known input–output pairs until the weights converge to within a specified tolerance.

The training can be very time consuming. In this study, a conjugate–gradient method was used to speed up the training[10]. In the conjugate–gradient algorithm, the gradient of the error with respect to the weights was calculated globally for all the patterns. The weights were then updated using a variable step size. This shortened the training time considerably. The compensation algorithm using the feed-forward network is described in the next section.

4.0 Electromagnetic Compensation Using Neural Network Computing

A neural network algorithm for emulating the CLS compensation method is outlined below. There are two approaches to training neural networks to compute CLS compensation excitations. The first approach would be to train up a multilayer feed-forward network with a very large number of permutations and combinations of surface error profiles with varying amplitudes. This method "covers all the bases" but it suffers from having no organized training plan. It would also be difficult to measure the networks performance to determine the number and types of training patterns required for the network to learn the CLS operation. It would in all likelihood be very large and training would probably be very time-consuming.

A second approach attempts to overcome these difficulties by decomposing the surface error profile into a truncated Fourier series approximation. Working with this approximation, the compensation can be obtained through superposition of a well-defined, finite set of trained networks. The advantage of this method is that the training is systematic and organized. It is easier to evaluate the performance of this system and make corrections to it. The formulation of this organized neural network electromagnetic compensation algorithm is given below.

The physical optics/aperture integration (PO/AI) formulation was decomposed into an approximate superposition of radiation integrals. In this initial study, infinite cylindrical reflectors were used. This simplifies the mathematical complexities and yields greater insight by reducing some of the numerical difficulties. Figure 3 shows the reflector geometry. The feed array elements are assumed to be y-polarized. The PO/AI solution for the radiated electric field from the antenna in Fig. 3 is given by[9]

$$E = \frac{-jk\eta}{2\pi\sqrt{r}}\, e^{-jkr}\, T(\theta) \qquad (2)$$

and $T(\theta)$ is[9]

$$T(\theta) = \int_{Aperture} [2N \times H']\, e^{jkr'\cdot\hat{r}} ds' \qquad (3)$$

where N is the surface normal, H' is the magnetic field at the reflector surface due to the feed array, r' is the vector that locates the points of integration on the surface, and $\hat{r}$ is the unit vector in the direction of observation.

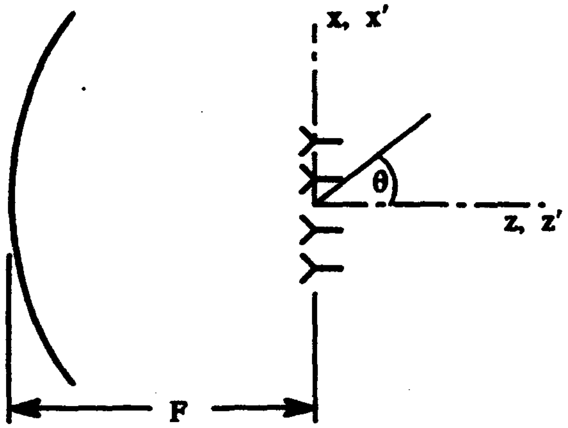


Figure 3. Geometry for the infinite cylindrical reflector.

For a distorted reflector it is desired to approximate the surface as a smooth parabola plus a finite series of perturbation functions, $\sum_i \Delta_i(x')$. In this study, a truncated Fourier series was chosen to approximate the surface error profile. The deviation from a smooth parabola is given by

$$\sum_i \Delta_i(x') = \sum_{m=1}^{M} C_{ma} \cos\left(\frac{2\pi C_{mb} x'}{D}\right) + \sum_{n=1}^{M} S_{na} \sin\left(\frac{2\pi S_{nb} x'}{D}\right) \tag{4}$$

where C_{ma}, S_{na} are the amplitudes of the Fourier terms, C_{mb}, S_{nb} are the number of periods of the cosine or sine across the reflector diameter D. After considerable manipulation, the integral from (3) for a distorted reflector with small errors can be be approximated by the decomposition[9]

$$T(\theta) \approx \sum_{m=1}^{M} T_{cm}(\theta) + \sum_{n=1}^{N} T_{sn}(\theta) - (M + N - 1)T_o(\theta) \tag{5}$$

where $T_{cm}(\theta)$ is equivalent to the integral from (3) for a reflector with a cosine error of magnitude C_{ma} and C_{mb} periods across the diameter, $T_{sn}(\theta)$ is equivalent to the integral from (3) for a reflector with a sine error of magnitude S_{na} and S_{nb} periods across the diameter, and $T_o(\theta)$ is equivalent to the integral from (3) for a smooth reflector. Therefore, given the Fourier coefficients of the surface distortion profile, the compensation excitations for a distorted reflector may be determined as a sum of compensation excitations for the individual reflectors with cosine and sine errors using the radiation integral decomposition from (5).

It is this decomposition which allows for the organized neural network compensation training algorithm. The CLS method is used to compute the training data (compensation excitations) for sets of reflectors with cosine and sine errors corresponding to the terms in (5). The CLS data are used to train sets of neural networks to interpolate the CLS excitations given the Fourier coefficients. A diagram illustrating this procedure is shown in Fig. 4. Each network consists of 1 input node that corresponds to the Fourier coefficient for a given sine or cosine error. The number of outputs for a network is twice the number of feed elements; half are for the real parts of the excitations and half are for the imaginary parts of the excitations. For each network in this study, there was one hidden layer.

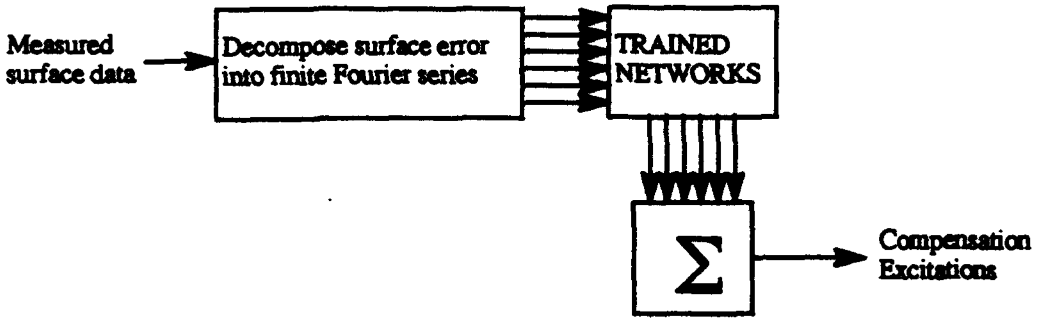


Figure 4. *Diagram illustrating the multilayer feed-forward (MLF) network implementation of the CLS method using the radiation integral decomposition method.*[9]

5.0 Compensation Results

The first example is used to illustrate the validity of the approximations in (5). The reflector geometry is that of Fig. 3. The focal length to diameter ratio is $F/D=1.0$ with a diameter of 100λ. The feed array has 13 elements spaced 1.0λ apart. The elements are assumed to have a "$\cos^q(\theta)$"-type pattern with $q=8$. The distortion profile is given by[9]

$$\Delta = (0.03\lambda)\cos\left(\frac{2(2\pi)x'}{100\lambda}\right) + (0.02\lambda)\sin\left(\frac{2(2\pi)x'}{100\lambda}\right) \qquad (6)$$

which corresponds to cosine and sine errors with 2 cycles across the diameter. Figure 5 shows the radiation pattern for the smooth reflector, distorted reflector without compensation, and the distorted reflector with the actual CLS compensation (no approximation). Figure 6 shows the comparison between the performance of the actual CLS compensation to that achieved by computing the CLS excitations for individual reflectors distorted by the cosine and sine from (6) and superposing these individual excitations using the approximation given in (5). The performance of the methods in Fig. 6 is almost identical.

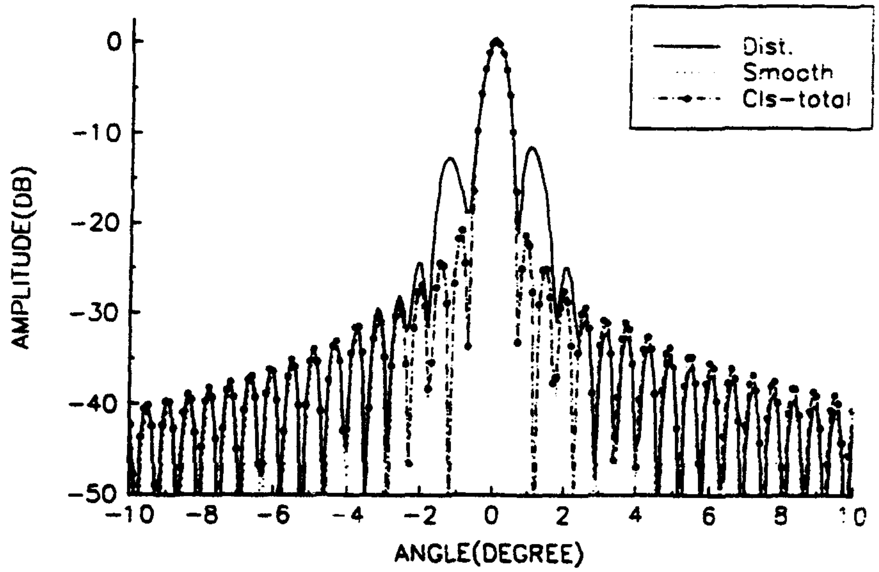


Figure 5. ***Radiation patterns calculated for the smooth reflector (dotted), distorted reflector using (6), without compensation (solid), and with actual CLS compensation – no approximations (circles).***[9]

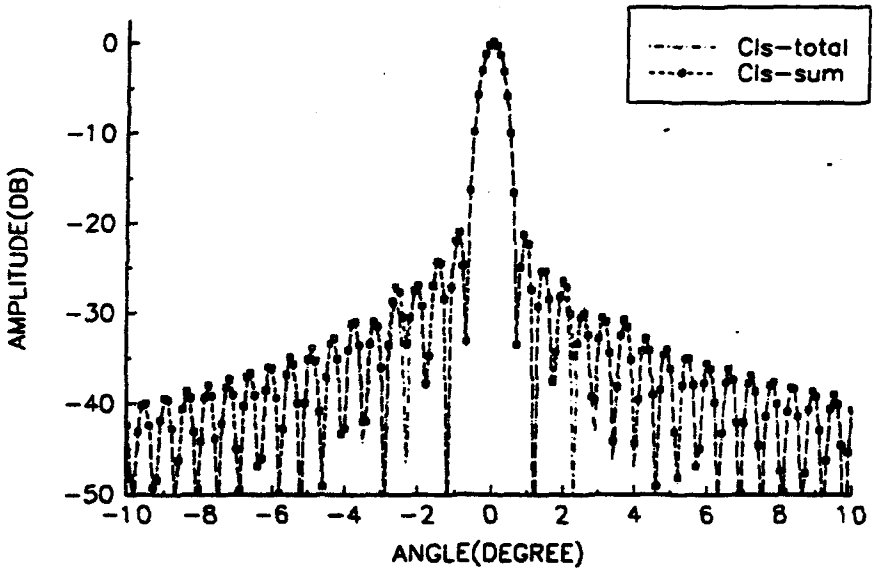


Figure 6. ***Comparison of the performance of the actual CLS compensation method – no approximation (dot-dash) and the superposition of individual CLS excitations (dash-square).***[9]

The second example illustrates the neural network compensation technique using the decomposition in (5). The reflector is the same as in the first example with the exception of the surface error profile which is now[9]

$$
\begin{aligned}
\Delta = {} & (0.015\lambda)\cos\left(\frac{(2\pi)x'}{100\lambda}\right) + (0.022\lambda)\cos\left(\frac{2(2\pi)x'}{100\lambda}\right) + \\
& (0.027\lambda)\sin\left(\frac{(2\pi)x'}{100\lambda}\right) + (0.011\lambda)\sin\left(\frac{3(2\pi)x'}{100\lambda}\right) \;.
\end{aligned} \tag{7}
$$

This Δ has multiple spatial frequency errors with 1, 2 and 3 periods across the reflector diameter. The error not only causes high side lobes but also scans the main beam off boresight.

The neural network compensation is accomplished through a set of 10 trained networks representing the first five integer spatial frequencies of both sine and cosine type errors. As noted above, the input to a network is the amplitude of the Fourier component. There are 40 nodes in the hidden layer and 26 outputs for each network. Figure 7 shows the radiation patterns for the smooth and uncompensated distorted reflector. Figure 8 shows the radiation patterns due to a distorted reflector without compensation, with the original CLS compensation, and with the neural network superposition. Figure 9 compares the original CLS solution and the neural network solution to the desired smooth reflector pattern. A few of the near-in side lobes for the neural network method are slightly higher than those for the original CLS compensation. This is partly due to the approximations and assumptions made in formulating the decomposition of the radiation integral (5) and partly due to the neural network simulation of the CLS compensation technique.

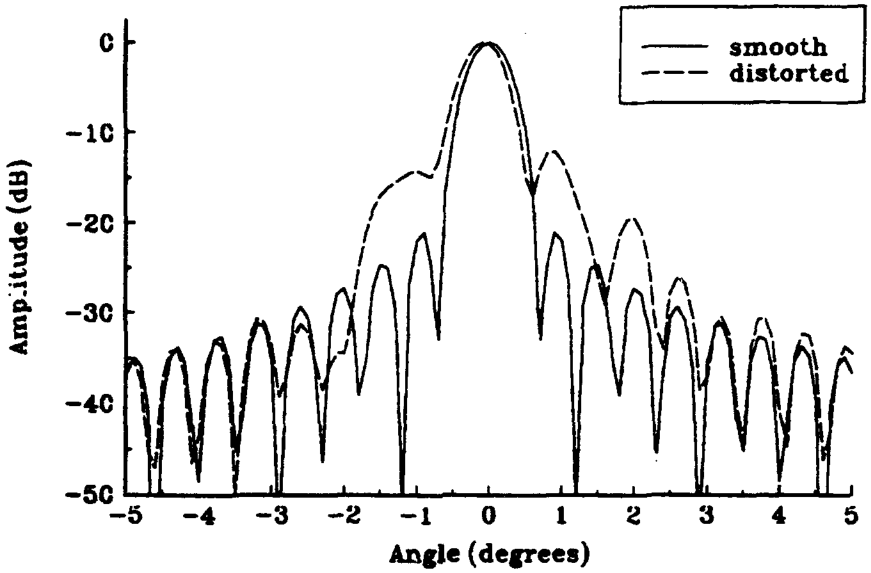


Figure 7. Radiation patterns for the smooth reflector (solid) and a distorted reflector without compensation.[9]

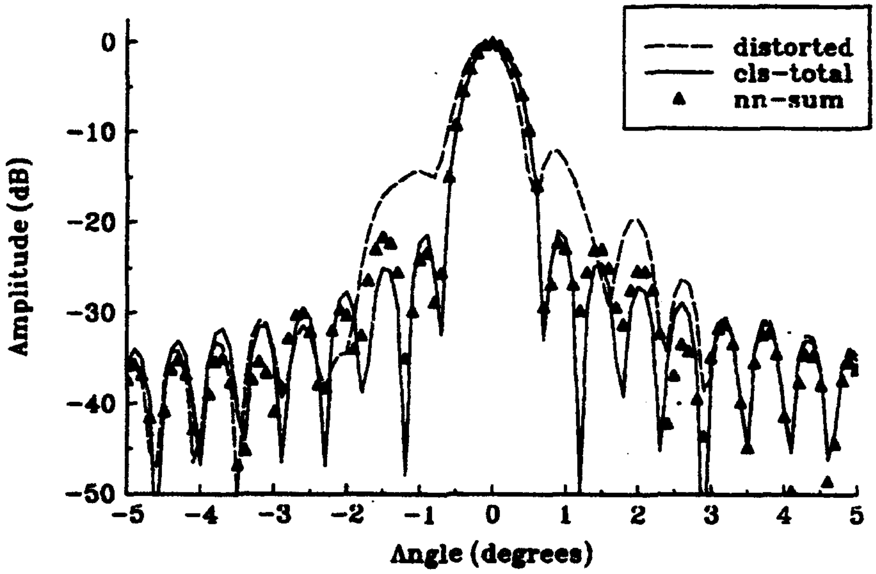


Figure 8. Radiation patterns for a distorted reflector for no compensation (dashed), original CLS compensation (solid), and neural network compensation (triangles).[9]

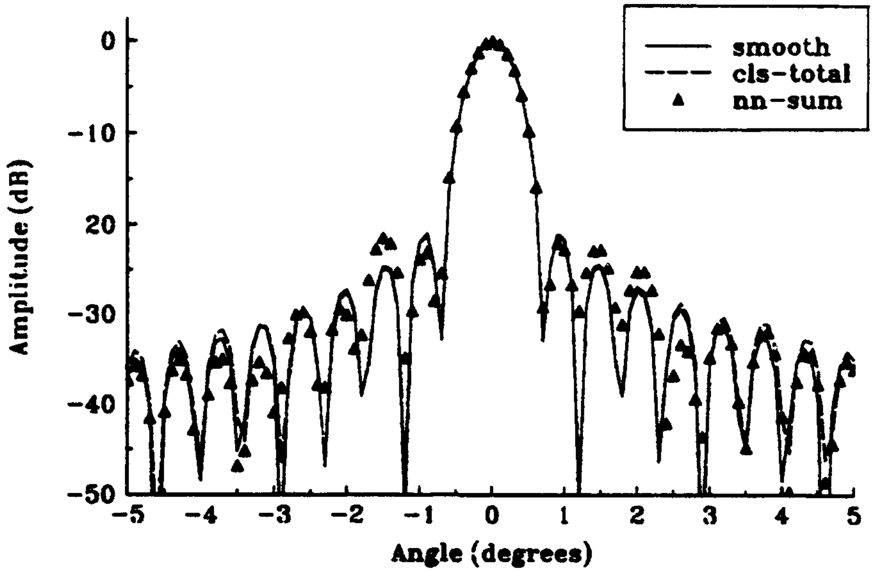


Figure 9. Radiation patterns for a smooth reflector (solid), distorted reflector with original CLS compensation (dashed) and neural network compensation (triangles).[9]

While there is slight degradation in the compensation performance, a significant advantage is gained in computational speed. Once trained, the neural network solution is much faster than the original CLS method. Table 1 shows a typical comparison of solution times for the cylindrical reflectors. The improvement in speed should be much greater when working with paraboloidal reflectors. The original CLS algorithm can take several hours to produce the compensation excitations while the neural network algorithm will continue to take just a few seconds.

Table 1. Computational time requirements for the infinite cylindrical reflector compensation problem from Example 2. The calculations were performed on a Sun SPARCstation 2.

Original CLS method	20 minutes, 31 seconds
Neural network compensation method	< 2 seconds
Training time for neural networks	5 to 10 minutes per network

6.0 Summary

This study analyzed an existing electromagnetic surface error compensation technique, CLS, and a new neural network formulation. The multilayer feed-forward neural network was investigated for its ability to approximate the performance of the CLS compensation technique. The radiation integral decomposition allowed for an organized, systematic way to train the set of networks and provided the means for easy expansion and modification.

The results showed that the neural network algorithm provided nearly equivalent performance compared to that of the original CLS method. The advantage of using the neural network algorithm is its computational efficiency. The training can be time consuming but is performed up front. Once trained, the neural network method is very fast and makes near real-time control possible.

References

1. Bailey, M. C. (1991) Determination of array feed excitation to improve performance of distorted or scanned reflector antenna, 1991 IEEE Antennas and Propagation Society Symposium Digest, London, Ontario: 175-178.
2. Smith, W. T. and Bailey, M. C. (1992) Quantization sensitivity of the constrained least squares compensation method, 1992 IEEE Antennas and Propagation Society Symposium Digest, Chicago, Illinois: 32-35.
3. Smith, W. T. and Bastian, R. J. (1993) Electromagnetic surface error compensation for reflector antennas using neural network computing, 1993 IEEE Antennas and Propagation Society Symposium Digest, Ann Arbor, Michigan: 750-753.
4. Rudge, A. W. and Davies, D. E. (1970) Electronically controllable primary feed for profile-error compensation of large parabolic reflectors, Proc. IEE, 117:351-358.

5. Dragone, C. and Hogg, D. C. (1963) Wide–angle radiation due to rough phase fronts, The Bell System Technical Journal, 42.

6. Blank, S. J. and Imbriale, W. A. (1988) Array feed synthesis for correction of reflector distortion and vernier beam steering, IEEE Trans. Antennas Propagat., AP–36:1351–1358.

7. Bailey, M. C., Cockrell, C. R. and Staton, L. D. (1989) Electronic compensation for reflector surface distortion to improve radiation pattern characteristics of antennas, NASA Technical Memorandum, 100652.

8. Pao, Y. (1989) Adaptive Pattern Recognition and Neural Networks, Addison Wesley Publishing Company, Reading, Massachusetts.

9. Bastian, R. J. (1993) Electromagnetic Surface Error Compensation for Reflector Antennas Using Neural Network Computing, Master's Thesis, University of Kentucky.

10. Charalambous, C. (1992) Conjugate gradient algorithm for efficient training of artificial neural networks, IEE Proceedings–G, 139:301–310.

THEORETICAL AND EXPERIMENTAL STUDIES OF MICROSTRIP REFLECTARRAYS USED FOR MOBILE AND SATELLITE COMMUNICATIONS

John Litva, Yuan Zhuang and Chen Wu
Communications Research Laboratory, McMaster University
Hamilton, Ontario, Canada L8S 4K1

[ABSTRACT] : The theoretical and experimental studies of microstrip reflectarray, for the applications in mobile satellite communications, is presented in this paper. A highly efficient full-wave analysis technique - combination of CG-FFT method and the complex discrete image technique - has been developed for the analysis of reflectarrays. Based on the technique, a design scheme is proposed and applied to some trial designs. A few prototypes have been designed, analyzed, constructed and tested. Experimental and theoretical results are compared and found to be in good agreement.

1 . INTRODUCTION

Mobile and satellite communication systems require high gain antennas which are compact in size and light weight. Downsizing antenna technology recently received a lot of attention recently by antenna researchers. The microstrip reflectarray (MRA in short) has been suggested as a possible antenna for mobile and satellite communications. [1]

The use of reflector antennas for beam forming is analogous to reflectors used in optics - the reflector surface is properly designed and fabricated to compensate for the phase differences of the fields travelling from the feed to different points on the reflector. A coherent phase-front for the reflected field is thereby generated. To reach the same goal, the microstrip reflectarray uses microstrip patches (see Fig.1), connected to open or short transmission lines to form a directional beam. The patches serve as re-radiators, and the connected microstrip transmission lines (we call them tails) as phase shifters. The tails are adjusted so as to generate a coherent field in a wanted direction.

Following the introduction of the concept of microstrip reflectarrays in 1987, only a few feasibility studies have been carried out. They have been primarily experimental in nature [2]. This is due to the fact that, as of yet, virtually no design tools have been developed for these types of structures. To add to this challenge, it should be noted that the structures are usually fairly large. In this paper, we present theoretical and experimental studies of MRAs. This work is carried out by using full-wave analysis techniques developed in our laboratory, which are a unique combination of the CG-FFT method with the complex discrete image technique. This technique is very efficient in terms of computer memory and computing time. It allows us to deal with very large MRAs (up to 1000 λ^2 area of 10 grids/λ) without any difficulty. Based on this full-wave analysis technique, we have proposed an optimized design scheme. Our design approach is based on studies of different shaped re-radiators, different element layouts, phase shift vs. length of microstrip tails, current distributions, and phase errors and tolerance analysis. We have designed, analyzed, constructed and tested some prototype arrays. The experimental results verify our theoretical studies.

2. ANALYSIS

To accurately analyze a very large microstrip reflectarray, in particular, " to accurately predict the backscaterred fields from thousands of non-uniformly illuminated microstrip patches with un-equal lengths of microstrip transmission lines, a very complex analysis technique needs to be developed [1]. " The Conjugate-Gradient Fast Fourier Transform Method (CG-FFT) is very suitable for analyzing large microwave structures [3]. When combined with the discrete image technique [4], CG-FFT becomes a very robust technique for analyzing very large microstrip antenna arrays [5]. In this paper, this technique is utilized to simulate a large microstrip reflectarray, as well as to predict the phase shifts of single patches with different lengths of tails, all in an array environment.

The general integral equation describing a microstrip antenna can be written as:

$$-\hat{n} \times \vec{E}^{inc}(\hat{r}) = \hat{n} \times \int_{s} \bar{\bar{G}}(\hat{r}, \hat{r}') \cdot \vec{J}(\hat{r}')dS \tag{1}$$

with the conducting surface denoted by S, the current $\vec{J}$, the incident field $\vec{E}^{inc}$, the unit vector normal to the surface $\hat{n}$ and the dyadic Green's function $\bar{\bar{G}}$. If the equation is written into an operator equation form, then: $\vec{E} = L(\vec{J})$. The iterations of CG algorithm minimize the functional $F(\vec{J}) =< \vec{r}, \vec{r} >= ||\vec{r}||^2$ where $\vec{r} = L\vec{J} - \vec{E}$. Because the integration in (1) is convolutional, an FFT can be used to compute the operation $L\vec{V}$ and $L^*\vec{V}$ (L^*:adjoint operator;$\vec{V}$: $\vec{J}$ or $\vec{r}$). The CG-FFT method is very efficient in terms of computer memory ($O(N)$) and computing time ($O(4N(1 + log_2N)$).

To overcome the difficulties of Sommerfeld-type Green's function related with microstrip structures, the full-wave discrete image technique is adopted into CG-FFT. This technique can convert the complicated Sommerfeld-type Green's function into a series of simple free-space-type Green's functions without losing any full-wave information. The computing time required for the generation of the discrete image is negligible compared to the CG-FFT procedure and for a given substrate (single or multilayer) only one set of image has to be found regardless of the antenna's configuration and layout.

In the simulation, the incident field $\vec{E}^{inc}$ is defined by the characteristics of the feed antenna. When the type and the location of the feed is decided, it is not difficult to calculate the incident fields precisely (magnitudes, phases and glazing angles) at every position within the reflectarray. After the calculation of the current distribution, $\vec{J}(\hat{r}')$ on the conducting surface, the Fourier Transform of the $\vec{J}$ on a single patch with tail can be carried out to give the phase information of the re-radiated field from the patch.

$$I_\xi(\theta,\phi) = \int dx' \int dy' \vec{J}_\xi(x',y') e^{-j\vec{k}_s\cdot\hat{r}'} \qquad (2)$$

with $\hat{r}' = x'\hat{x} + y'\hat{y} + z'\hat{z}, \xi = x$ or y and $\vec{k}_s = k(sin\theta cos\phi\hat{x} + sin\theta sin\phi\hat{y} + cos\theta\hat{z})$.

3. DESIGN SCHEME AND RESULTS

Based on the simulation approach described above, combined with conventional optics, reflector antenna and array theories. an optimized design scheme is proposed , which is shown explicitly in diagram form in Fig.2. It should be noted that a number of different types of reflector antennas can be designed. These include the parabolic, the elliptical, hyperbolic, circular, triangular corner, and, of course, flat sheet. The feed can have a big offset from the center of the array, as long as the phase shifts from each patch are predicted accurately before designing the layout of the complete array.

Following the scheme given in Fig.2, we have simulated , designed and tested some trial X-band linear microstrip reflectarrays, with different numbers of patches. Each patch is resonant at 12 GHz. The ϵ_r of the substrate used is 2.2, and its thickness is 1.59mm. Fig.3 shows the experimental set-up in our anechoic chamber.

First, the relationship between the phase shift of the reflected field from a single patch and the length of the connected tail is investigated. This is presented by the solid line in Fig.4. To show the effect of mutual coupling, a single patch in an array environment has been analyzed, and the phase shift relation is shown in Fig.4(dashed line). To study the effect of different shaped patches, the phase shift of patches with indented tails are also simulated(Fig.4, dotted line).

The whole array is analyzed using CG-FFT. In the simulation, we found that the current distributions on the patches are effected by the attached tails. This behavior is due to the impedance looking into the tail at the junction with the patch varying with the length of the tail. For instance, the impedance should be near zero when the tail length is $\frac{\lambda_g}{4}$. The current distribution on the patch for this case should be close to zero because of the short-circuiting effect of the tail. As the length either increase or decrease with respect to $\frac{\lambda_g}{4}$, the current increase monotonically. This is known as $\frac{\lambda_g}{4}$ effect. These relationships is shown in Fig.5.

The current distribution and radiation pattern for the first design example are shown in Fig.5. The radiation pattern is calculated from the current distribution. The H-plane (focused plane) pattern is shown in Fig.7(solid). The main beam is focused in the direction we designed. The measurements for the array are carried out in our anechoic chamber. The measured pattern is compared with the simulated pattern. Reasonably good agreement is shown. Due to the absence of a proper 12GHz low-noise-amplifier, the pattern measured outside of 50^0 is dominated by noise. We can find that the currents on the central patches are smaller than those on the side patches (V-shape) which results in a relatively high side-lobe-level(SLL). This is because of (1) the edge effect of the backscattering field from the arrays; (2) the improper arrangement of the length of tails (or say the location of the reference zero-degree patch).

To overcome the destructive tail effect and edge effect, we propose the following scheme.

- Add some dummy elements at both sides of the array. From Fig.8, we can find that if we put some patches with $\frac{\lambda_g}{4}$ tails at the both sides of the array (Range B), the short-circuit effect and the edge effect can cancell each other.

- Combine two ways of tunning. In Dr.Pozar's work [?], they find that the phase of the re-radiated field can be shifted by changing the length of the patch. So we propose that by combining the two techniques for tunning the patches, i.e., by changing the length of tails and/or the length of patches - we can achieve an optimal current distribution as well as the desired phase shifts.

Fig.9 & 10 give an example of a MRA before and after the implementation of the above optimal adjustment. We can find the improvement of the SLL of the modified MRA with a little cost in the width of the main beam.

Fig.11 & 12 give the measured results for two other MRAs without using the above scheme. Both of them focus at the designed direction accurately, 0^0 and 12^0 respectively. The work of design, construction and test of a large 2D MRA with optimal adjustment is under the way.

4. CONCLUSIONS

From the simulations and measurements carried out in this paper, the following conclusions can be drawn. To achieve accurate focusing of the microstrip reflectarray, one of the key points is to determine the precise phase shift of the re-radiated signal of each single patch in the array. This phase is controlled caused by the length of the connected short transmission lines. Intuitively, a linear relationship between the phase shift and the length of the tails is expected. This can be proven true if the equivalent transmission line model is used. But at a higher frequency (e.g. X-band) and in an array environment, the tail cannot simply be treated as a pure transmission line. The effects of re-radiation, surface waves and mutual coupling between neighbouring patches and tails , etc., have to be considered. This is true in the analysis of complete array as well as in the prediction of the phase change of each patch. Our proposed approach was taken all these factors into consideration. The simulation results show that the phase shifts vs. lengths of tails is not a simple linear relation. By carrying out a study of different shaped patches, we find that we can make the array more compact and achieve better focusing by properly choosing the shape and layout of the patches in the array. AAs well, by carefully arranging the patches with zero phase shift, by putting some dummy elements at the edge of the array and by using two tunning techniques, we can improve the performance of MRA (gain, SLL, etc.). Continuing research is under way.

References

[1] John Huang " Microstrip Reflectarray " IEEE AP-S Int. Symp. Dig., 1991, pp.612-615

[2] Chang,D.C.,Huang,M.C. : " Microstrip Reflectarray Antenna with Offset Feed " Electronic Letters, 1992, pp.1489-1491

[3] T.K.Sakar, E. Arvas & S.S.Rao : " Application of FFT and Conjugate Gradient Method ... " IEEE Trans. AP, 1986, AP-34, pp635-640

[4] D.G.Fang, et.al.,"Discrete image theory for horizontal electric dipoles in a multilayer medium" IEE Proc.Part H, Vol.135, PP297-303, 1988

[5] Yuan Zhuang, Keli Wu, John Litva : " Application of CG-FFT Combined with Discrete Image Technique in the Analysis of Microstrip Antennas" IEEE AP-S & URSI Joint Symp. Dig., 1992

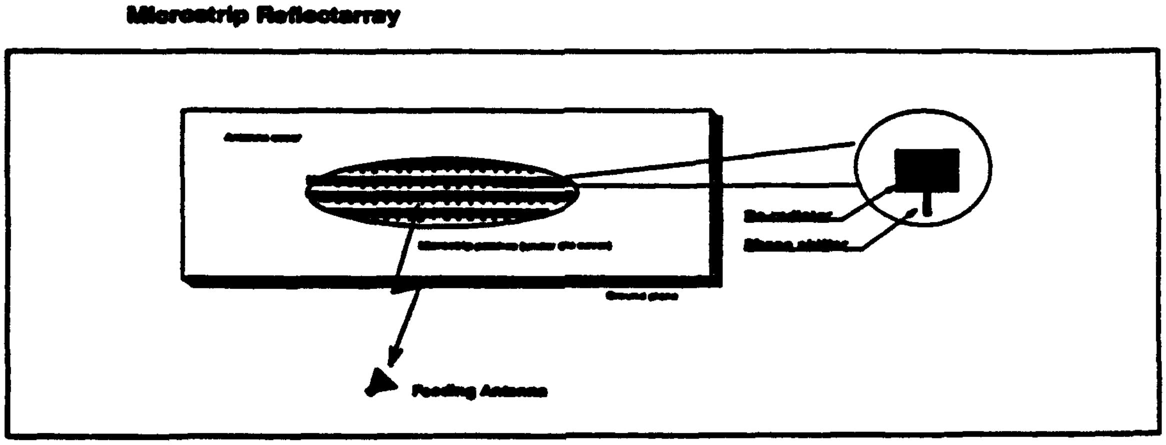

Microstrip Reflectarray
Antenna array
Microstrip patches (under the curve)
Ground plane
Re-radiator
Phase shifter
Feeding Antenna

FIGURE 1

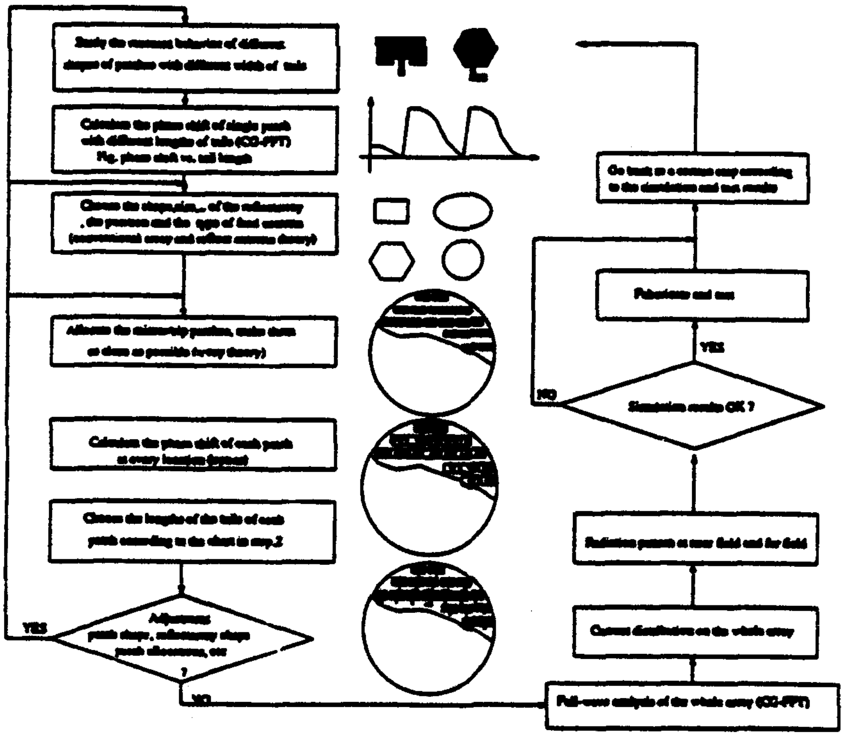

Study the resonant behavior of different shapes of patches with different width of tails
Calculate the phase shift of single patch with different lengths of tails (CG-FFT)
Fig. phase shift vs. tail length
Choose the shape,size,.. of the reflectarray, the position and the type of feed antenna (conventional array and reflector antenna theory)
Allocate the microstrip patches, make them as close as possible (array theory)
Calculate the phase shift of each patch at every location (layout)
Choose the lengths of the tails of each patch according to the chart in step.2
YES
Adjustment patch shape, reflectarray shape patch allocation, etc ?
NO
Full-wave analysis of the whole array (CG-FFT)
Current distribution on the whole array
Radiation pattern at near field and far field
Simulation results OK ?
NO
YES
Fabricate and test
Go back to a certain step according to the simulation and test results

Diagram of microstrip reflectarray design scheme

FIGURE 2

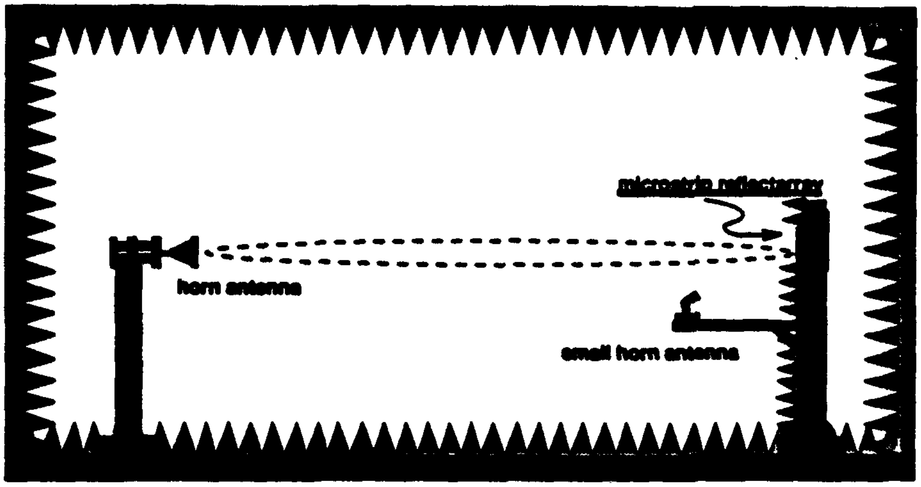

microstrip reflectarray
horn antenna
small horn antenna

Fig.3 Experiment set-up for microstrip reflectarray measurement in an anechoic chamber

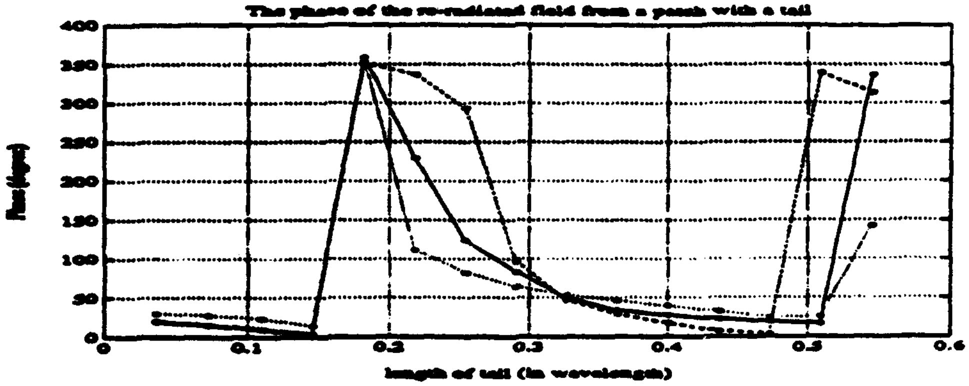


Fig.4 The phase shift of the re-radiated field from a patch with a tail
Solid line : single patch;
dashed line: patch in array environment
dot line : patch with indented tail

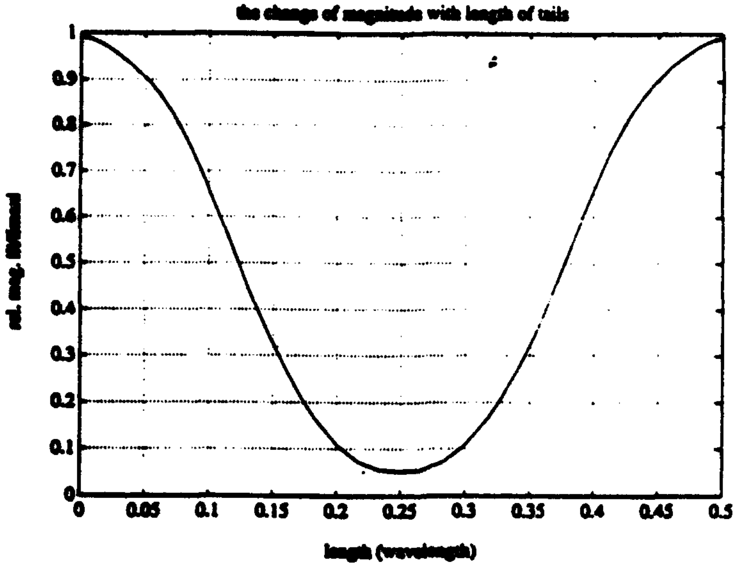


Fig.5 The relationship between the current magnitude and the length of tail of a reactangular patch.

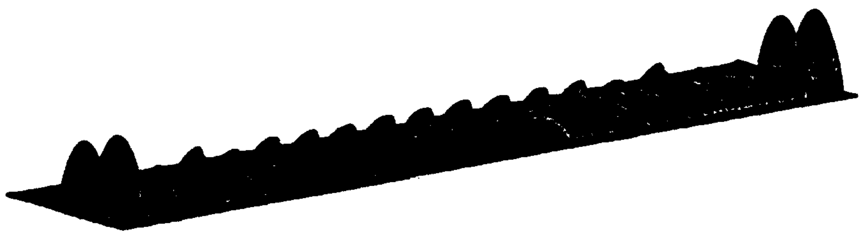


Fig.6 The current distribution on a microstrip reflectarray (on some centeral patches)

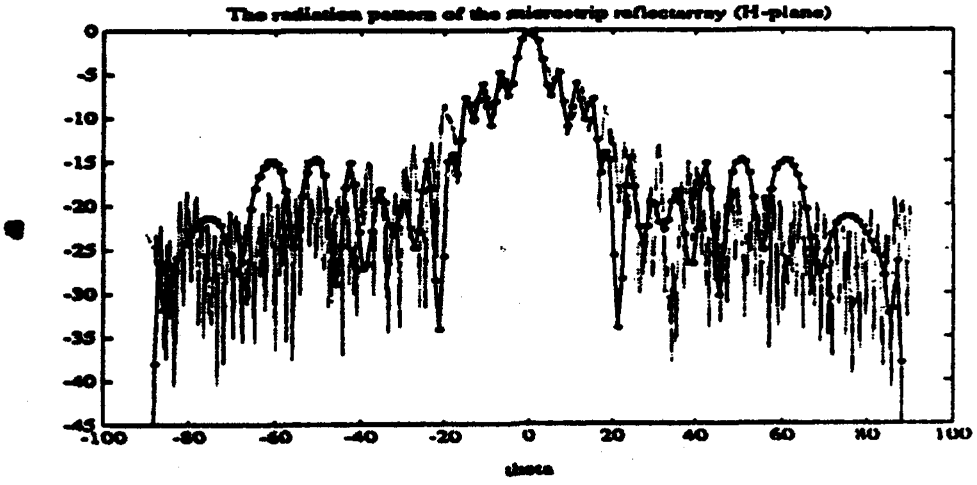


Fig.7 The radiation pattern of the microstrip reflectarray
solid line : simulation result dor line : measured result

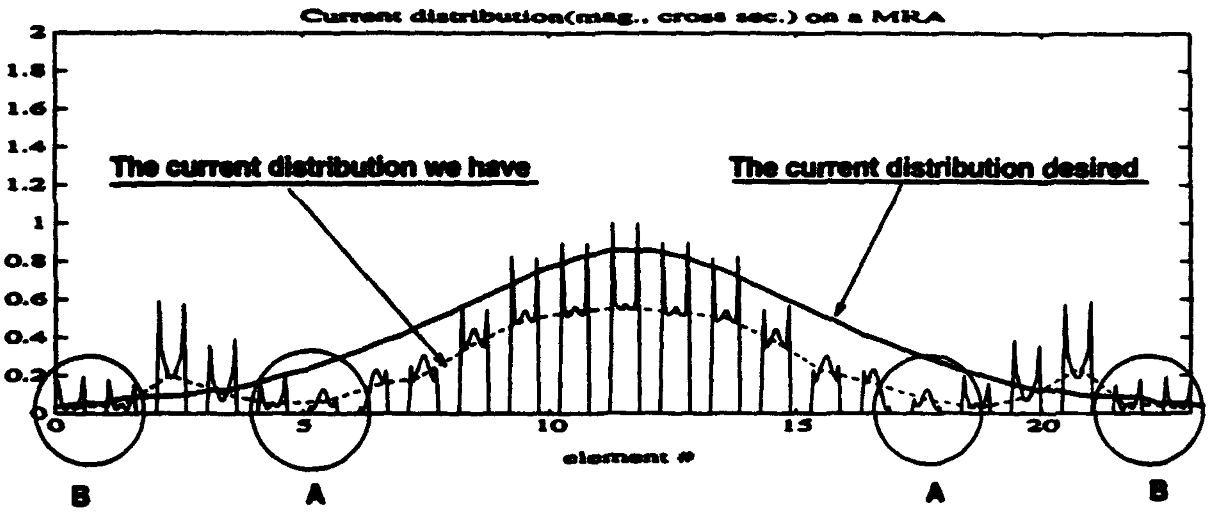
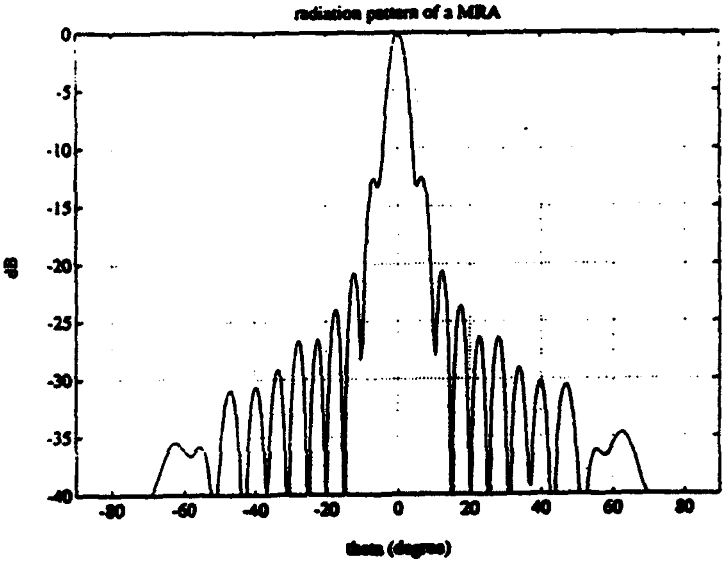


Fig.8 The magnitude of current distribution (cross section) on a MRA

Range A : two-way-tunning range

Range B : dummy elements range



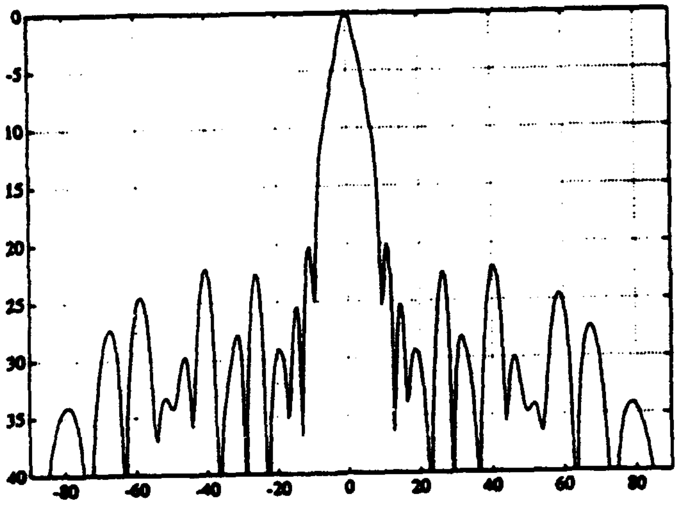


Fig.9 Radiation pattern of the MRA without two-way-tunning adjustment

Fig.10 Radiation pattern of the MRA with two-way-tunning adjustment

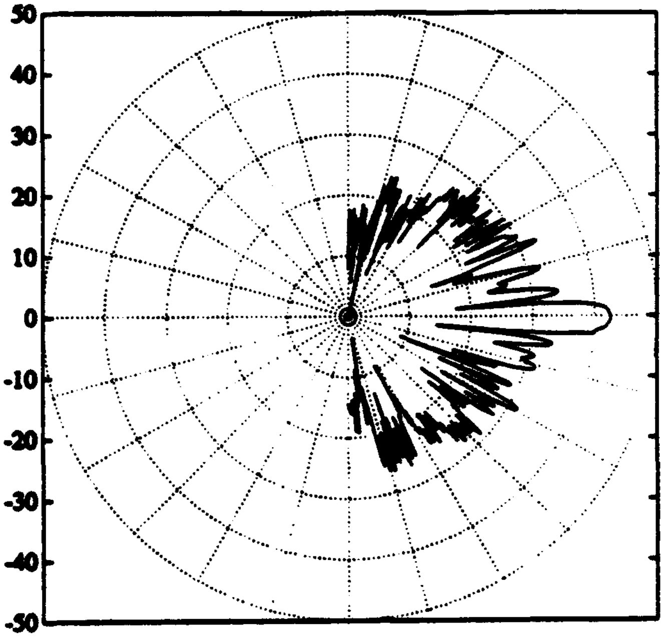


Fig.11 Measured radiation pattern of a linear microstrip reflectarray with 0 degree main beam direction.

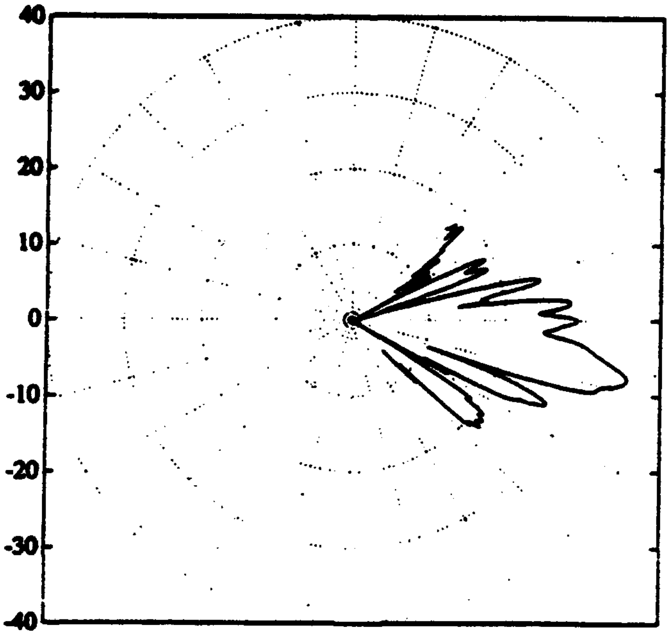


Fig.12 Measured radiation pattern of a linear microstrip reflectarray with 12 degree main beam direction

A STUDY OF THE SYNTHESIS OF A LARGE COMMUNICATIONS APERTURE USING SMALL ANTENNAS

Tom Cwik, Vahraz Jamnejad, and George Resch
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

ABSTRACT

The Jet Propulsion Laboratory is currently engaged in a study to develop a quantitative understanding of the performance, cost, and technical risks associated with synthesizing a large aperture from an array of smaller aperture antennas. The array will be a receive-only system, operating simultaneously at S-band and X-band. This Small Aperture Array Study will parameterize costs of the entire array as a function of the antenna element diameter for a prescribed G/T (gain divided by system noise temperature). As a benchmark, the prescribed G/T will be that of a small number of Deep Space Network 70m antennas (one to three). In this paper, the costs for the antenna subsystem are parameterized. The entire system cost parameterization is available elsewhere.

1. INTRODUCTION

The Jet Propulsion Laboratory is currently engaged in a study to develop a quantitative understanding of the performance, cost, and technical risks associated with synthesizing a large aperture from an array of smaller aperture antennas. Such an array would support the communications links to spacecraft engaged in planetary and solar system exploration. This study represents the conceptual exploration of a particular evolutionary path that is open to the Deep Space network. The array will be a receive only system, operating simultaneously at S-band and X-band. The product of the study is an analytic model that relates the total system cost to the diameter of the elemental apertures for a given G/T (i.e., total antenna gain divided by total system temperature). As a benchmark, the prescribed G/T will be that of a small number of Deep Space Network 70m antennas (one to three). Costs for the complete system will be parameterized. These include the antennas, radio and intermediate frequency amplification, signal distribution, combiner electronics, and the monitor and control needed to operate the array in a synchronous fashion. This paper documents the results of the antenna subsystem cost analysis.

2. ANTENNA COST MODEL

The antenna system is an obvious and major component in the overall array cost model. As will be detailed, the antenna system will be divided into subsystems that include all mechanical and structural components, the foundation, and microwave optics (including the feed system, but not including any electronic packages). To simplify the cost estimation process and keep it within a limited time and budget, "off the shelf" technology is to be used for each subsystem. It was decided to contract to two companies specializing in antenna ground station design and fabrication so that detailed antenna subsystem costs could be supplied, and that the estimated costs would not be speculative. These two companies are

TIW Systems, Inc., Sunnyvale, CA, and Scientific Atlanta, Inc., Atlanta, GA (SA). The two companies have previously supplied JPL with antenna systems, and therefore are familiar with the specific requirements and procedures of the DSN. Specific tasks that the contractors were to complete are as follows:

- For eight antenna diameters ranging from 3 to 35 meters, production techniques will be investigated and a preferred design for each antenna subsystem will be specified.
- The design will include specifying antenna optics for each antenna size based on cost, manufacturability, and performance.
- Each subsystem will be further divided into non-recurring and recurring costs.
- Because of the large number of antennas that could be fabricated (especially at the smaller diameters), it is expected that an economy of scale will be encountered. This cost study should outline breakpoints in production where costs drop for a given diameter as more antennas are fabricated.
- To assist in the probabilistic determination of the number of antennas needed to maintain a prescribed G/T margin, the cost estimates should outline antenna components which critically affect reliability, and detail the costs of critical components as a function of reliability.

Due to the limited time and budget of the contracts, the last three items were not examined in great detail. The costs in this section therefore do not reflect any reductions that may be gained by mass production of antenna systems specifically designed for this DSN array application.

3. ANTENNA SPECIFICATIONS

The number of antennas needed to synthesize the G/T of a 70m antenna is a function of the diameter and system noise temperature of the antennas. Shown in Table 1 is the range of the number of antennas needed for each of the eight diameters specified to the contractors. The column of minimum units corresponds to cooled amplifiers and enough antenna elements to comprise one station, while the maximum number of units corresponds to uncooled amplifiers and enough elements to comprise three stations. This range was specified to allow for economies of scale in production methods to surface, and for a complete parameterization of the antenna-amplifier system based on system noise

Table 1. Minimum and Maximum Antenna Elements

	Units	
Diameter (m)	Minimum	Maximum
3	545	27,000
5	196	10,000
10	49	2,500
15	22	1,100
20	12	615
25	8	394
30	5	274
35	4	201

temperature and antenna diameter. Common sense dictates that an array of inexpensive 3-meter antennas using expensive cooled amplifiers, as well as expensive 35-meter antennas using inexpensive uncooled amplifiers, should produce extremes in the cost model. These extremes would be expected to bound the cost model.

The antenna optics are broken into two regimes. For small diameter antennas, a frequency selective subreflector is used to separate S-band—arranged as a prime focus system—from X-band, which is arranged in a Cassegrain system. For larger diameter antennas, both bands operate in a Cassegrain system, with the bands separated by either a dual-frequency (concentric) feed, or a frequency-selective surface (FSS) diplexor. It was expected that the break would occur in the range of 10–20-meter antenna diameters. This breakpoint option and frequency-combining method were left to the contractor. TIW arrived at designs which used prime focus S-band designs, including an FSS subreflector, for diameters up to and including 10 meters, and Cassegrain configurations with a dual-frequency feed for diameters of 15 meters and larger. Scientific Atlanta arrived at similar designs but with a breakpoint where the dual-frequency feed is used for diameters greater than 21 meters.

To gain a richer understanding of the antenna system cost model, the antenna was broken into eight subsystems. These are as follows:

- Antenna Support Structure. Designs for all antenna sizes were conventional elevation over azimuth configurations. Due to the range of antenna sizes considered, modifications based on production, shipping, and assembly were made to arrive at a final design.

- Main Reflector Surface. Again, based on antenna diameter, different panel production methods were used in the final design.
- Axis Drive. Includes actuators, drive gearboxes, and bearings.
- Position Control. Includes encoders, motors, cabling, and controls.
- Feed System (Including FSS). As noted above, different feed systems were used at the diameter breakpoints specified by the contractor.
- Foundation. No below ground enclosure supplied.
- Power Supply. Includes distribution on site.
- Shipping, Installation, and Testing. Different strategies for installation and testing were used based on antenna diameter.

Summaries of the designs as well as the cost information are contained in the final reports supplied by the contractors.

4. PERFORMANCE REQUIREMENTS

The performance requirements specified to the contractors are those contained in the JPL DSN Document 810-5, Volume 1: Existing DSN Capabilities. The specifications necessary for this study are listed in Table 2.

5. THE ANTENNA COST MODEL

Traditionally, antenna cost models have followed a power law

$$C = a + bD_E^x \qquad (1)$$

Table 2. Antenna Element Specifications

Parameter	Specification
Operating Frequency	From S-band to X-band
Axis Coverage:	
Elevation	0° to 90°
Azimuth	±200°
Reflector Surface	Solid aluminum
Environments:	
Precision Operation:	
Wind	10 mph gusting to 12 mph
Rain	2 inches per hour
Temperature	0°F to 115°F
Normal Operation:	
Wind	30 mph gusting to 36 mph
Rain	2 inches per hour
Temperature	0°F to 115°F
Survival:	
Wind	100 mph (stowed)
Seismic	0.3 G horizontal and 0.15 G vertical
Hail	Up to 1-inch-diameter stones
Temperature	–20°F to 180°F
Drive-to-Stow	60 mph
Maximum Tracking Rates:	
Velocity	0.4°/sec
Acceleration	0.4°/sec^2
Maximum Slew Rates:	
Velocity	0.4°/sec
Acceleration	0.2°/sec^2
Site Location	Australia
Soil Conditions	3,000 psf bearing capacity at 3 feet below grade (no piles required)
Axis Configuration	Elevation over azimuth

Table 2. Antenna Element Specifications (Continued)

Parameter	Specification
Pointing Accuracy:	
Precision Operation	0.1 beamwidth
Normal Operation	0.2 beamwidth
Surface Accuracy:	
Precision Operation	0.030-inch RMS
Normal Operation	0.035-inch RMS
Concrete Foundation	Minimum height (no building room required)

where a represents a constant fixed cost, b is a constant, and D_E is the dish diameter. The exponent x is the critical parameter in the cost model, driving costs as the antenna size increases. This parameter has been previously estimated by examining costs of existing antennas and fitting the above power law to the data. One early estimate[1] gave x as 2.78, and this number has been widely quoted. In this study we will fit the above power law to the overall antenna element cost, but will also examine the subsystem costs using fits appropriate for the subsystem. For example, the feed subsystem does not have to increase with dish diameter, but may show step function breaks when changing from prime-focus S-band system to dual-frequency Cassegrain systems.

Figure 1 shows the cost estimates from the two contractors for the antenna elements as a function of diameter. Scientific Atlanta supplied data for more diameters than specified because they have existing systems or cost data at 13, 16, 18, 21, and 32 meters. The SA data is not as smooth as the TIW-supplied cost data

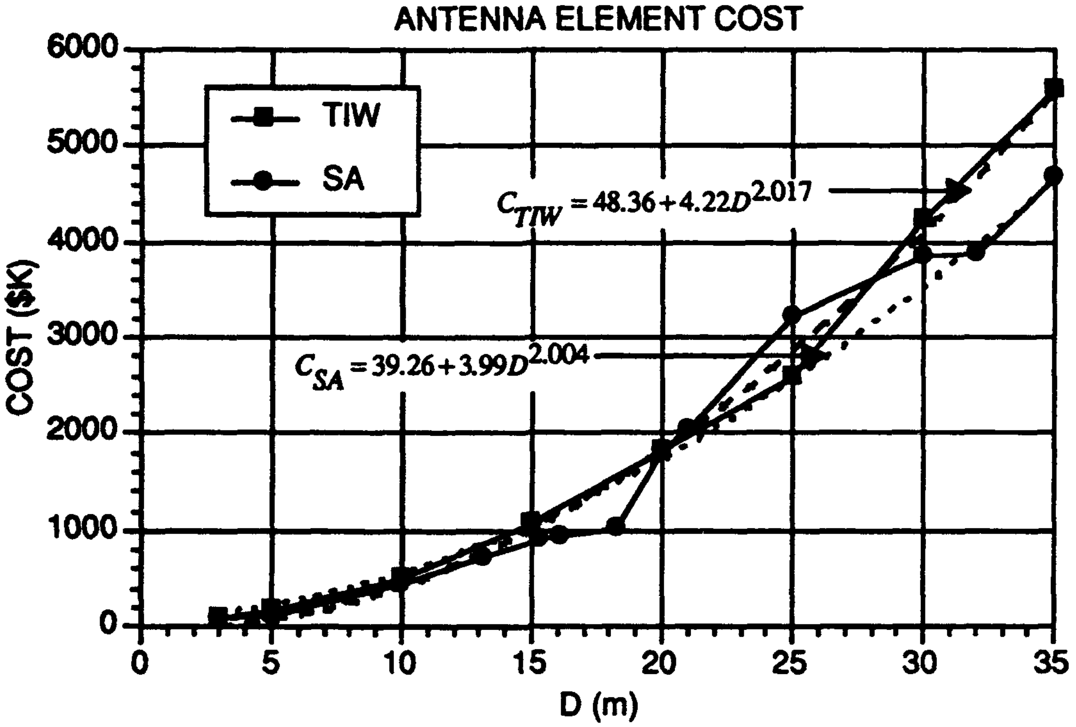


Figure 1. Total antenna element cost and power law fits to data. Costs in fits to data are in units of $K.

because of design variations at some diameters. Specifically, SA supplies an 18-meter system where the structure, foundation, and shipping, installation, and test subsystems are optimized for cost.

Power law fits to the data are also shown on the plot. For both data sets, it is seen that the cost increases as the diameter is *squared*, counter to higher powers previously published. It is interesting to note that the Project Cyclops study[2] came to the same conclusion for 25-meter to 150-meter antennas. The fact that antenna costs scale as diameter is squared profoundly affects the overall conclusions of this study.

6. ANTENNA SUBSYSTEM COSTS

Cost data for the eight antenna subsystems are presented in Figures 2–9. Costs for specific subsystems at specific antenna diameters are plotted on the charts for both contractors. The circle and square symbols denote the points where data was supplied, and solid lines connect the data points. On each chart, fits have been made to the data. Where appropriate, power law fits have been made; otherwise, step functions are used to model breaks in the data. For some components it is clear that the SA data does not have an obvious fit to a particular cost model. As mentioned previously, this is due to optimizations of certain diameters for cost reduction.

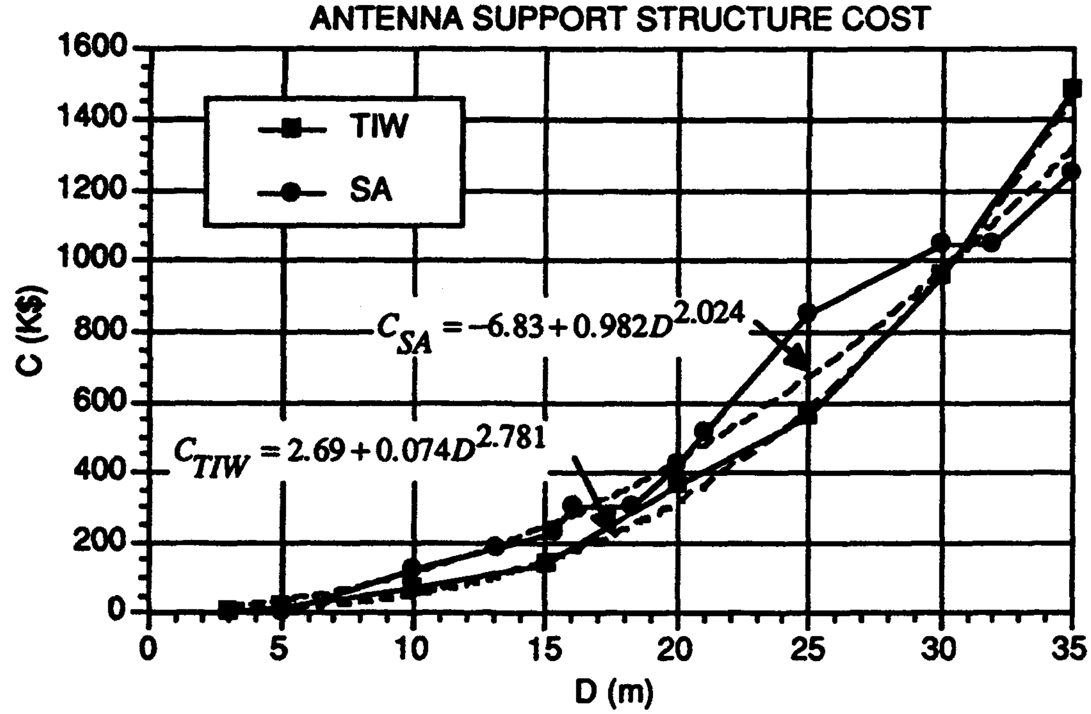


Figure 2. Cost and power law fits to data for antenna support structure. Costs in fits to data are in units of $K.

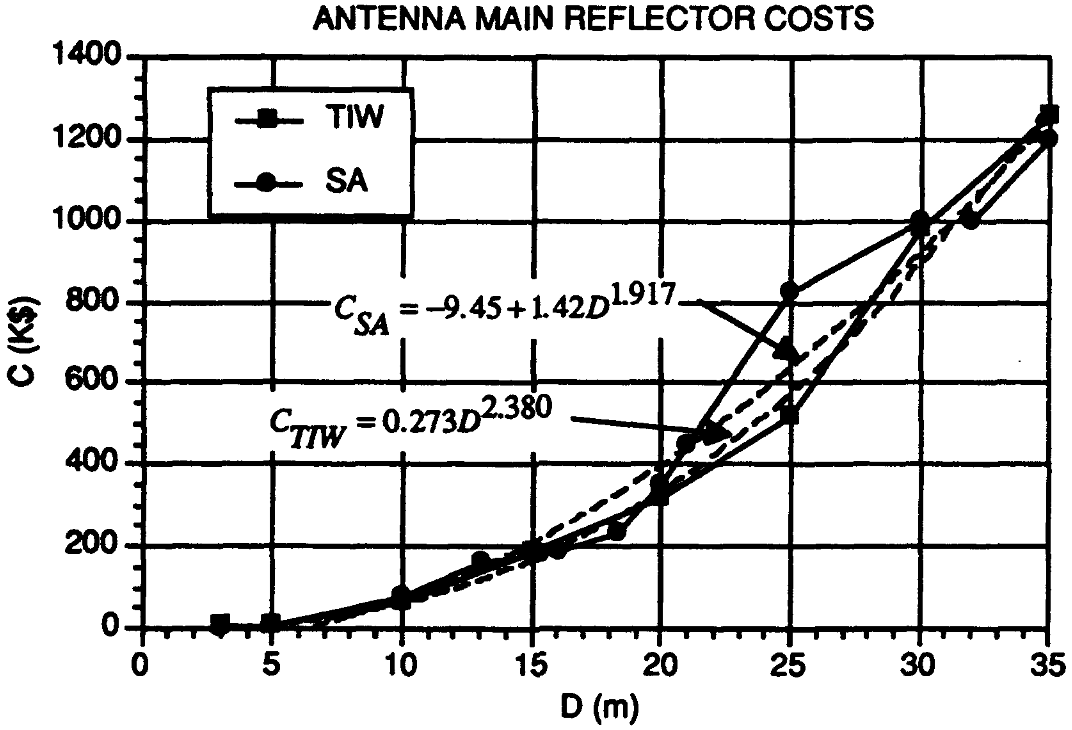


Figure 3. Cost and power law fits to data for antenna main reflector. Costs in fits to data are $K

7. SUMMARY

For the purposes of the overall array cost model, the cost power law fits from Figure 1 are sufficient to model the antenna system. When fit to a power law, the data from the two contractors are remarkably similar:

$$\begin{aligned} C_{TIW} &= 48.36 + 4.22D^{2.017} \quad (K\$) \\ C_{SA} &= 39.26 + 3.99D^{2.004} \quad (K\$). \end{aligned} \tag{2}$$

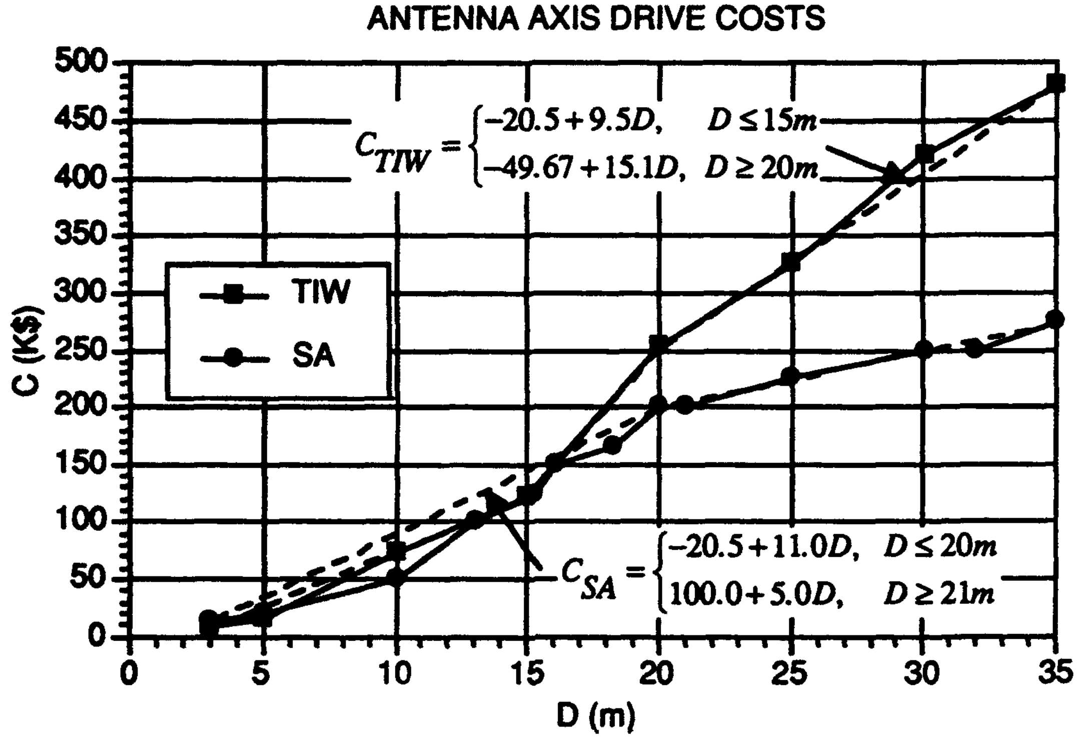


Figure 4. Cost and fits to data for antenna axis drive.
Costs in fits to data are in units of $K.

For better local fits to the data, or for individual subsystem cost data, the individual models shown in Figures 2–9 can be used. An antenna system cost model made up of the individual subsystems is then

$$C = C_{SUP} + C_{REF} + C_{AX} + C_{POS} + C_{FEED} + C_{FOUN} + C_{POW} + C_{SIT} \quad (K\$) \tag{3}$$

where the individual subsystem costs are given in the figures.

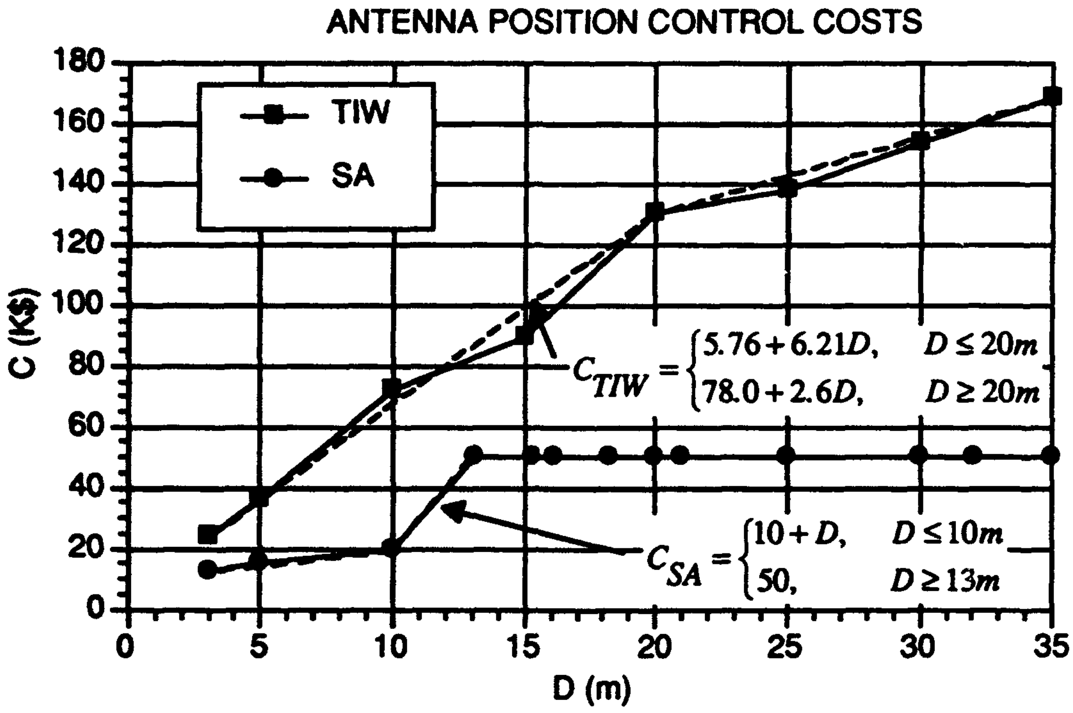


Figure 5. Cost and fits to data for antenna position control.
Costs in fits to data are in units of $K.

It is interesting to break down the costs by subsystem, examining the fraction each subsystem contributes to the total antenna cost, as well as the scaling of each subsystem. Figure 10 is the percent of total cost for each of the eight subsystems for the TIW data. It is seen that structure, reflector, and shipping, installation, and test subsytem costs increase with diameter size; feed, position control, and power subsystem costs decrease, while foundation and axis drive costs are relatively constant. For 3-meter antennas, the feed and position control subsystems contribute 57% of total cost—this is an obvious area for cost reductions for high quantity production. To extrapolate cost scaling for larger systems, the individual cost models (three) were calculated for diameters up to

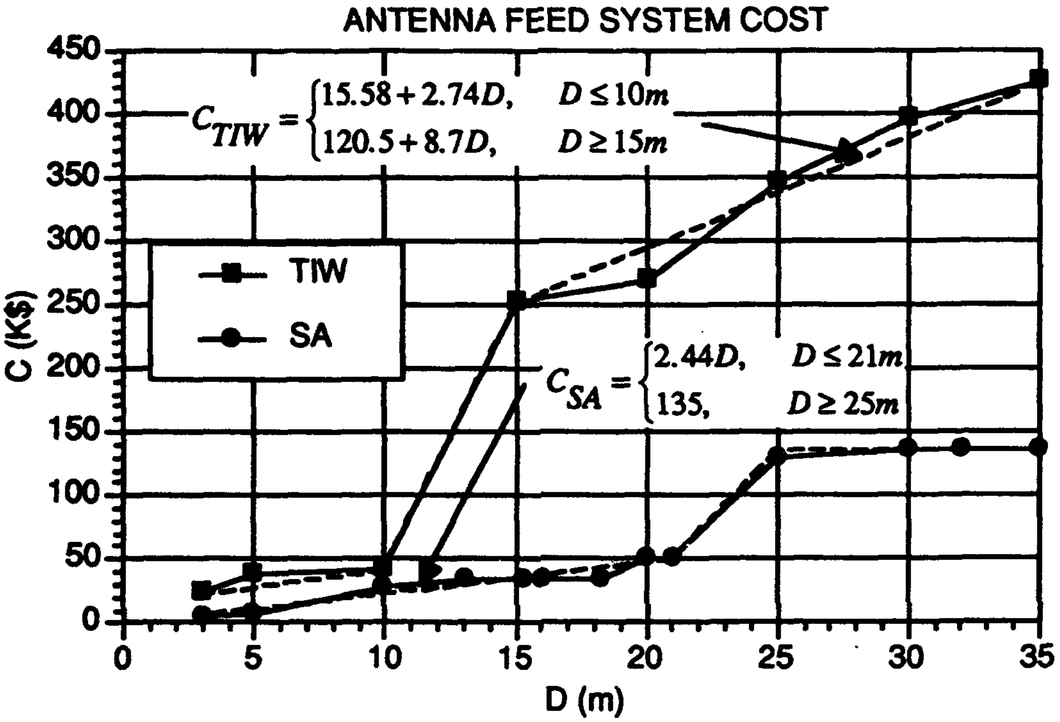


Figure 6. Cost and power law fits to data for antenna feed system. Costs in fits to data are $K.

100 meters. The costs were calculated based on the individual cost models for TIW data in Figures 2–9 with power law fits made to the resulting extrapolated data. It was seen that when a power law was fit to data up to 50 meters, the cost scaled as $D^{2.27}$; for fits to 70 meters the costs scaled as $D^{2.40}$; and for fits to 100 meters the costs scaled as $D^{2.50}$. These costs are, of course, extrapolations to the small antenna diameter data and are speculative.

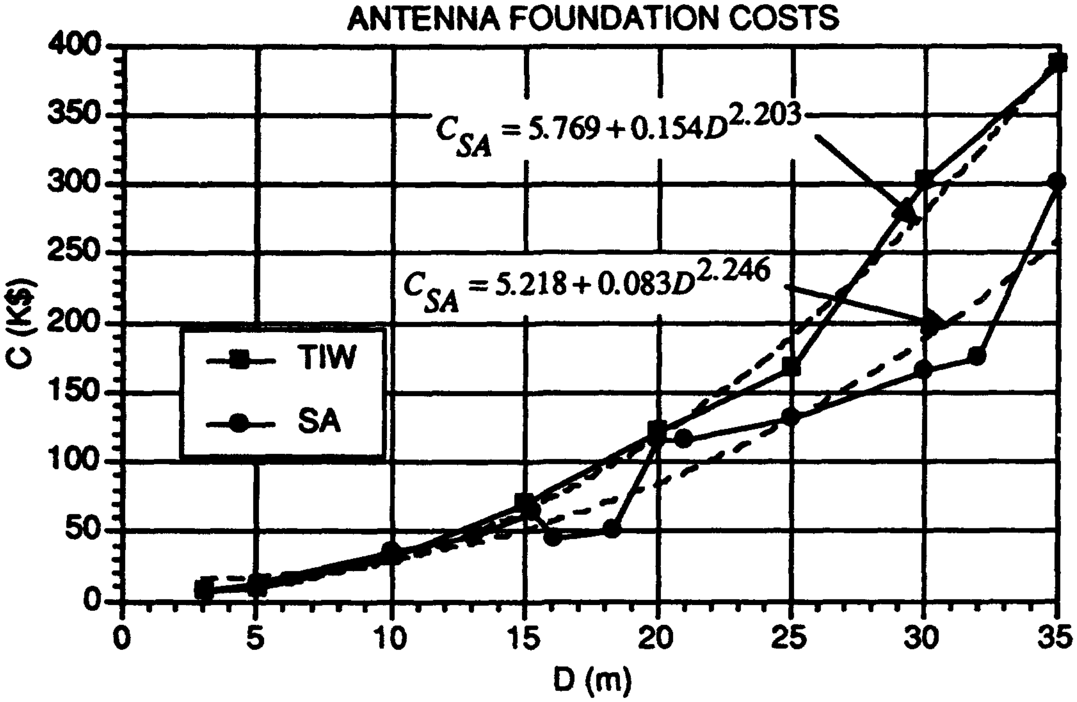


Figure 7. Cost and power law fits to data for antenna foundation subsystem. Costs in fits to data are in units of $K.

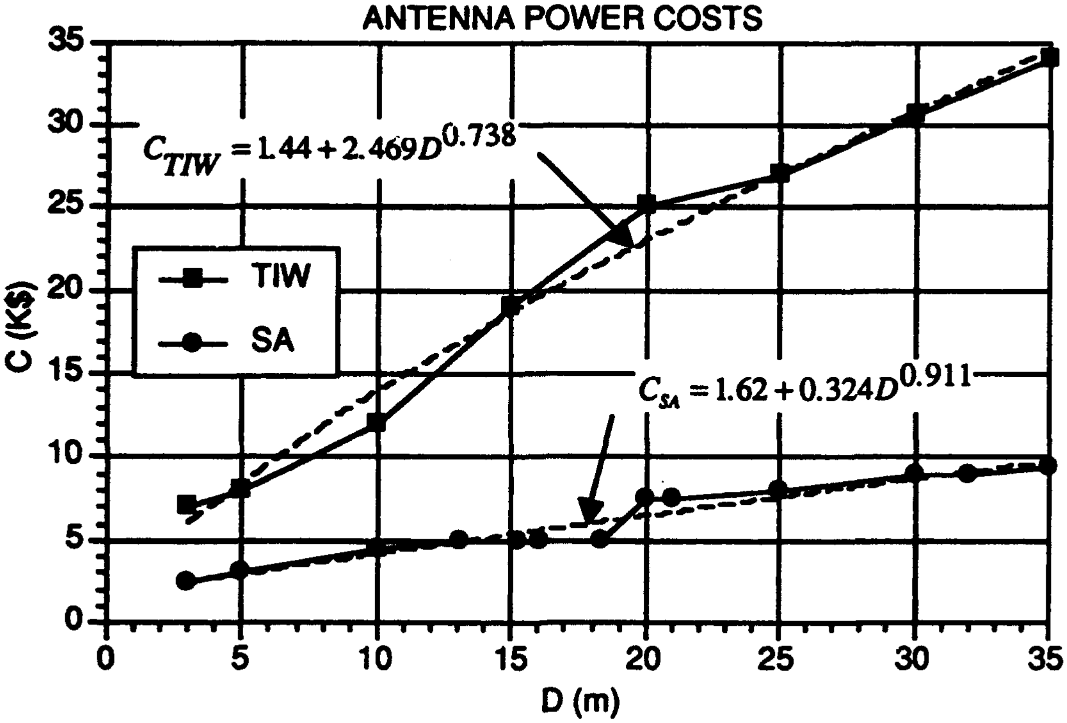


Figure 8. Cost and power law fits to data for antenna power subsystem. Costs in fits to data are in units of $K.

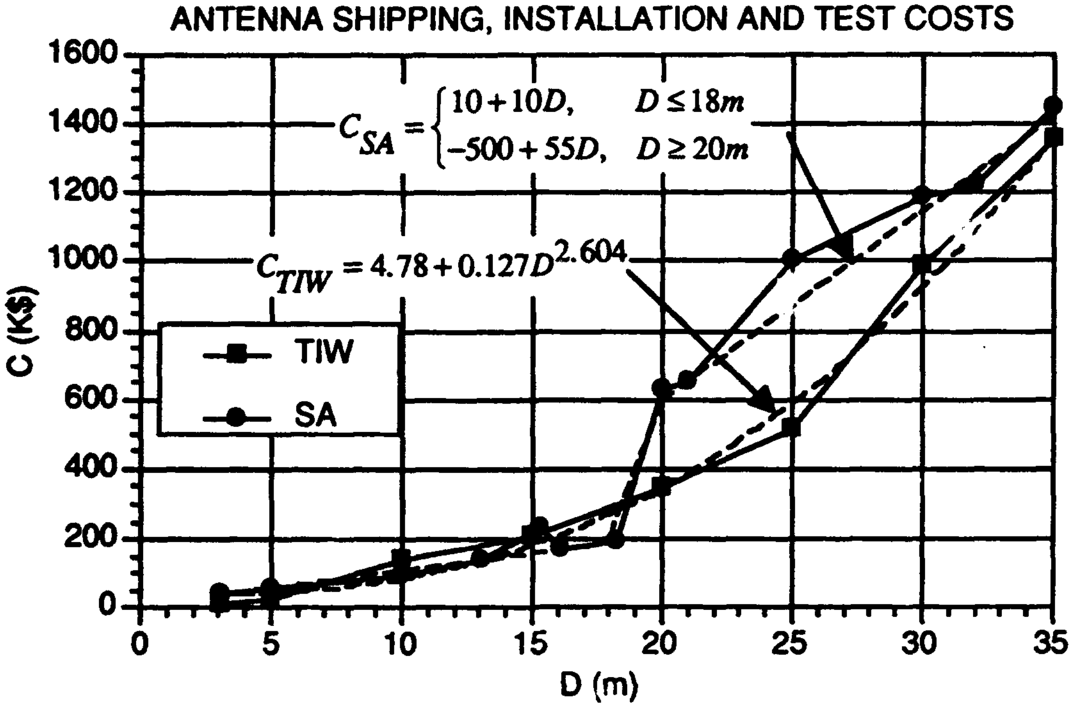


Figure 9. Cost and fits to data for antenna shipping, installation, and testing subsystem. Costs in fits to data are in units of $K.

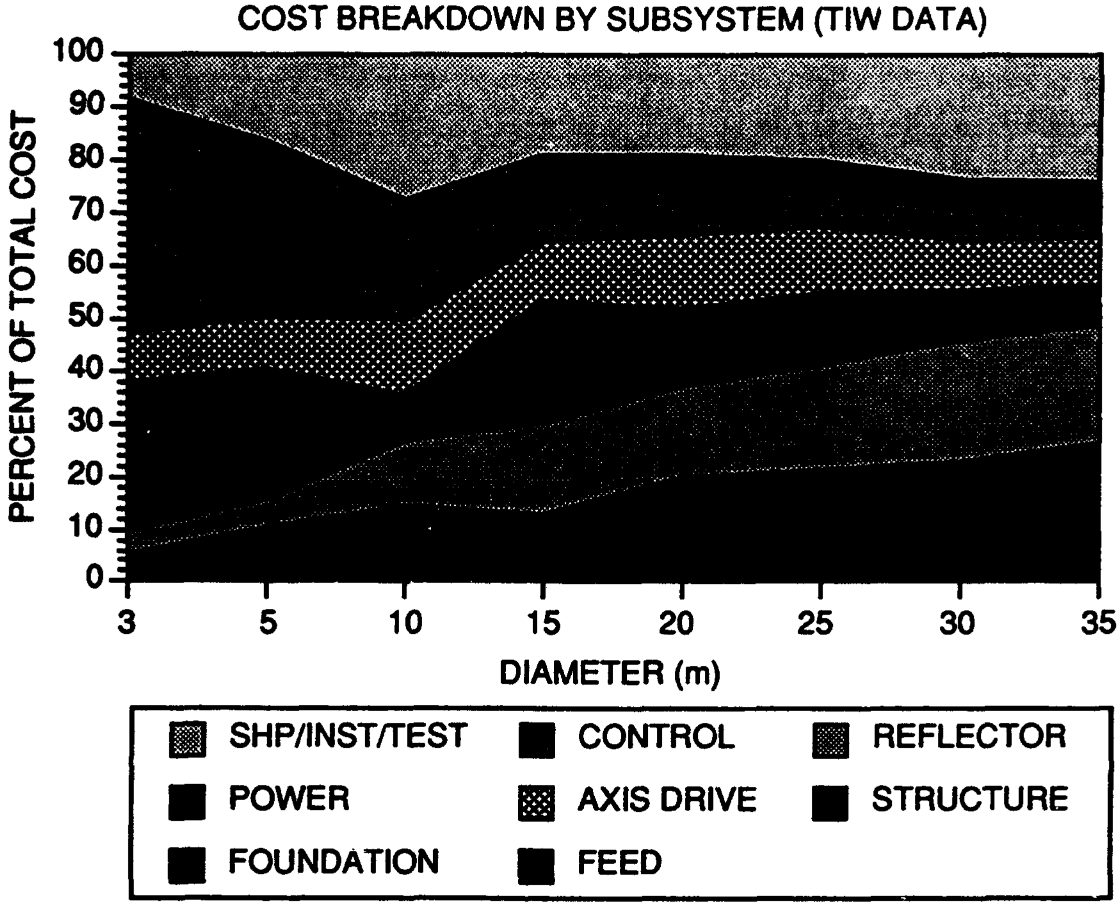


Figure 10. Cost breakdown by subsystem as fraction of total antenna cost.

8. ACKNOWLEDGMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

9. REFERENCES

1. Potter, P.D. (1965) Large Antenna Apertures and Arrays for Deep Space Communications. *JPL Technical Report No. 32-848.*
2. *Project Cyclops, A Design Study for Detecting Extraterrestrial Life*, CR 114445, NASA/Ames Research Center, 1971.

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESC Program Offices (POs) and other ESC elements to perform effective acquisition of C3I systems. In addition, Rome Laboratory's technology supports other AFMC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.